

**LOCALIZED SAFETY EVALUATION OF  
PERMANENT VARIABLE MESSAGE SIGNS (VMS)**

**A Thesis  
SUBMITTED TO THE FACULTY OF  
UNIVERSITY OF MINNESOTA  
BY**

**Ning Zhang**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
MASTER OF SCIENCE**

**Gary A Davis, Adviser**

**August 2014**

© Ning Zhang 2014

## **Acknowledgements**

First, I would like to thank my academic advisor, Dr. Gary A. Davis, for all his support and contribution to my thesis. He is very responsible, patient, and helpful reviewing my paper, and providing valuable guidance.

In addition, I would like to thank Dr. John Hourdos, and Dr. Kathleen Harder, for their professional advices on my thesis.

Then, I would like to thank Paul Morris from SRF Consulting Group, John McClellan and Kristina Krepela from the Minnesota Department of Transportation, for sharing the data, and responding to my technical questions.

Finally, I would like to thank my family and friends for all their kindly help and support.

## **Abstract**

The Variable Message Sign (VMS) has been deployed in Minnesota since 1960s. The evaluations for the effectiveness of the signs are critical for their appropriate installation and deployment. In this study, five VMS devices were selected along Interstate-94 between downtown Minneapolis and St. Paul. The data collection period ranged from January 2006 to December 2012. First, a linear model was used to investigate the effect of VMS on the speed changes, in the corresponding impact regions. The analysis was conducted using two scenarios with eight different conditions. The results revealed that the speed changes, which were influenced by the display of VMS messages, were within 2.0 miles per hour (mph) for all the conditions. Moreover, adverse weather, and times of day were unlikely to affect the model results. Second, a  $2 \times 2$  contingency table, and a logistic regression model were used to explore the association between the deployment of VMS and the crash occurrence. Odds ratios for the probabilities of crashes under the impact of VMS were estimated. The results indicated that the deployment of VMS messages was not likely a risk factor for crash occurrence. The estimated odds ratios for both warning and informative message types were not significantly different from 1.0.

## Table of Contents

List of Tables .....	v
List of Figures .....	vii
Chapter 1: Introduction .....	1
1.1 The Deployment of VMS in Minnesota.....	1
1.2 Brief Summary of Existing VMS Evaluations.....	2
1.3 Safety Evaluation Methodologies in this Study.....	4
1.4 Organization of the Thesis .....	5
Chapter 2: Literature Review .....	6
2.1 Evaluations of Variable Message Signs.....	6
2.1.1 Qualitative Evaluation .....	6
2.1.2 Quantitative Evaluation .....	9
2.2 Studies on the Crash-Prone Traffic Conditions .....	13
2.2.1 Traffic Speed and Crashes .....	14
2.2.2 Congestion Levels and Crashes .....	15
2.2.3 Density (Occupancy) and Crashes .....	16
2.3 Impact of Adverse Weather on Crashes.....	17
Chapter 3: Data Collection and Preparation .....	19
3.1 Definition of Study Scope.....	19
3.2 VMS Log Data Collection .....	20
3.3 Crash Data Collection .....	23
3.3.1 Crash Data Sources .....	23
3.3.2 Advantages and Limitations .....	23
3.3.3 Crash data Used in this Study .....	25
3.4 Traffic Data Collection .....	26
3.4.1 Crash Time Estimation .....	26
3.4.2 Traffic Conditions Prior to the Crashes .....	30
3.5 Weather Data Collection.....	33
3.5.1 Weather data Sources.....	33
3.5.2 Advantages and limitations.....	33
3.5.3 Weather Stations Selection and Data Validation .....	34

3.5.4 Weather Data Used in this Study .....	34
Chapter 4: Impact of VMS on Short-Time Change of Traffic Speed .....	37
4.1 Data Preparation.....	37
4.1.1 VMS Logs Data .....	37
4.1.2 Traffic Speed Data .....	38
4.1.3 Weather Data .....	39
4.1.4 Data Combination and Filtering.....	39
4.2 Model Construction and Methodology .....	42
4.3 Analysis of VMS Impact .....	45
4.3.1 The Impact of VMS in Adverse Weather Conditions.....	46
4.3.2 Analysis of Difference Scores .....	47
4.3.3 Analysis of Covariance .....	51
4.4 Discussion .....	54
Chapter 5: Case-Control Study .....	56
5.1 Data Preparation.....	56
5.1.1 VMS Logs Data .....	56
5.1.2 Crash Data.....	57
5.1.3 Non-crash Controls .....	58
5.2 Methodology and Model Construction .....	60
5.2.1 2×2 Contingency Table.....	60
5.2.2 Odds Ratio .....	61
5.2.3 Logistic Regression Model .....	63
5.3 The Deployment of VMS and Crash Occurrence .....	64
5.3.1 Unadjusted Odds Ratio .....	64
5.3.2 Adjusted Odds Ratio .....	66
5.4 Discussion .....	74
Chapter 6: Conclusion.....	76
6.1 Discussion of Methodologies and Findings .....	76
6.2 Achievements and Limitations .....	79
Bibliography .....	82

## List of Tables

TABLE 1 Detailed Information for Selected VMS .....	21
TABLE 2 Haghani's Grouping Table for VMS Message Types .....	22
TABLE 3 Detailed Information Provided by Different Crash Data Sources .....	24
TABLE 4 Number of Crashes in Each VMS Impact Region .....	25
TABLE 5 Crash Diagrams and Codes .....	25
TABLE 6 Traffic Metric list and Description.....	31
TABLE 7 Potential Weather Station Locations and Data Availability .....	35
TABLE 8 Weather Stations Used in this Study.....	35
TABLE 9 VMS Sites Selection for the Study of VMS Impact on Speed.....	38
TABLE 10 Scenarios for Testing the Change of Average Speed under Different Conditions .....	45
TABLE 11 Regression Results With and Without Weather Variables Comparison (Scenario One Example) .....	46
TABLE 12 Model Selections (Scenario One Example) .....	47
TABLE 13 Significance Test for the Average Speed Change in Scenario One .....	49
TABLE 14 Regression Result for Scenario Two.....	49
TABLE 15 Significance Test for the Average Speed Change in Scenario Two .....	50
TABLE 16 Comparison of Regression Results for Difference Scores Model and Analysis of Covariance Model (Scenario One Example) .....	52
TABLE 17 Cross-Classification of VMS Statuses and Crashes.....	61
TABLE 18 Joint and Marginal Probabilities of VMS Statuses and Crashes.....	61
TABLE 19 Distribution of Crashes and Non-crash Controls .....	64
TABLE 20 Odds Ratios and Hypothesis Test Results for Individual VMS Impact Regions .....	66
TABLE 21 Univariable Logistic Regression Results for Selected Traffic Metrics.....	67
TABLE 22 Possible Explanatory Variables for Logistic Model .....	68
TABLE 23 Regression Results for All the Selected Explanatory Variables (“Big Model”) .....	70

TABLE 24 Regression Results for “Small Model” .....	70
TABLE 25 Regression Results for “Medium Model” .....	71
TABLE 26 Logistic Regression Model Selection .....	71
TABLE 27 Regression Results for “Medium Model” with Aggregated VMS Message Type .....	73
TABLE 28 Hosmer-Lemeshow Tests Results for Medium Model .....	74



## List of Figures

FIGURE 1 Study Corridor and VMS Locations (Screenshot from Google Maps) .....	19
FIGURE 2 VMS Impact Region.....	20
FIGURE 3 The Frequencies for Each VMS Message Type .....	21
FIGURE 4 The Median Durations for Each VMS Message Type .....	22
FIGURE 5 Flow-density Model for VMS Site V94E08.....	27
FIGURE 6 Illustration of Crash Time Estimation .....	30
FIGURE 7 Potential Weather Stations Selected (Screenshot from Google Maps) .....	34
FIGURE 8 Shockwave Critical Distances to VMS Impact Regions .....	41
FIGURE 9 95% Confidence Intervals for the Average Speed Change .....	42
FIGURE 10 Frequency of Average Speed Change .....	43
FIGURE 11 Histogram of standardized residuals .....	51
FIGURE 12 Speed Curves for Different Times of Day.....	53
FIGURE 13 Speed Curves for Different VMS Statuses.....	53
FIGURE 14 Speed Change for Different Weather Conditions.....	54
FIGURE 15 Crash Frequencies by Time of Day (All 183 Crashes Collected from five VMS Sites).....	58
FIGURE 16 Crash Frequencies by Day of Week (All 183 Crashes Collected from five VMS Sites).....	59

## **Chapter 1: Introduction**

The purpose of this study is to evaluate the operational and safety effects of variable message sign (VMS). The evaluations mainly focused on the influence of VMS on the vehicle speed change, and the crash occurrence in reasonably specified impact regions.

### **1.1 The Deployment of VMS in Minnesota**

According to the Minnesota Department of Transportation (MnDOT), variable message sign (VMS) is defined as “a traffic control device whose message can be changed manually, electrically, mechanically, or electromechanically to provide motorists with information about traffic congestion, traffic crashes, maintenance operations, adverse weather conditions, roadway conditions, special events, or other highway features.” (MnDOT, 2012) In some publications, variable message sign is also named as changeable message sign (Harder et al., 2003), and dynamic message sign (Mounce, 2007). The history of the deployment of VMS in Minnesota dates back to 1960s. (Levinson and Huo, 2003) Currently, the signs are widely used on many major freeways and trunk highways, providing real-time guidance and traffic information to road users with various message types. The benefits of VMS are expected to improve the traffic operation, and to reduce the risk of non-recurring incidents, especially secondary crashes.

Permanent VMS and portable VMS are two types of signs that are generally used in practice. The former is installed on the ground or on other highway superstructure along the road, and the latter is usually moved with a truck or other portable vehicles that can be assigned to a required location (MnDOT, 2012). The overhead permanent VMS is of critical interest in this study, and it was abbreviated to VMS in the discussions below. For a VMS, the types of message that are authorized to be displayed include: incidents, work zones, travel times, adverse weather, road condition, special events, abducted child alerts, traffic safety campaigns, and test messages (MnDOT, 2012). The VMS analyzed in this study showed one message in three lines. Generally, the first line described an event, such as “CRASH” and “STALLED VEHICLE”; the second line described the location of the event, such as “AT HWY 280”; and the third line described the impact of the event or the guidance to drivers, such as “LANE CLOSED” and “REDUCE SPEED”.

Minnesota State Patrol, MnDOT Maintenance Dispatch, and MnDOT Freeway Operations officers have the access to the deployment of VMS. The activation is controlled by operators when a verifiable incident or non-recurring event affects or will affect the normal traffic flows (MnDOT, 2000). For incident management, a message turns on only after the event is observed and verified by RTMC operators, or requested by Minnesota Department of Public Safety (DPS) field personnel. For construction and maintenance activities, VMS is only deployed for short-term projects, or the first three days of long-term projects (MnDOT, 2000). If a long-time or repeated message is need, a permanent sign will be considered rather than a variable sign. The priorities of the messages typically follow from safety concerns, road (lane) closure, and congestion reminder. (MnDOT, 2000) Commonly, the messages are kept updated 24 hours a day, and 365 days a year.

## **1.2 Brief Summary of Existing VMS Evaluations**

The objective of VMS performance evaluations is to guarantee the devices are effectively operated, and the messages are properly displayed. In terms of the previous studies, the evaluations were conducted in two directions: qualitative evaluation and quantitative evaluation. Generally, qualitative evaluations were driver-survey oriented. Analyses based on survey results aimed at identifying drivers' "satisfaction" towards the deployment of VMS, and investigating drivers' possible reactions corresponding to the VMS messages (Tay, 2008). Quantitative evaluations were divided into two major groups: mobility evaluation and safety evaluation (Mounce, 2007). The former mainly focused on the impact of VMS on corridor traffic characteristics, such as travel time, total delay, speed patterns, diversion rate, and congestion; the latter emphasized on the potential impact on corridor crash rates, and the severity of crashes.

The types of safety evaluation can be summarized as "with-and-without" study, "on-and-off" study, "before-and-after" study, and simulation based study. Literally, "with-and-without" study compared the crash rates in numerous road sections with, and without the installation of VMS, and used statistical models to address the issue. Road sections were selected with similar geometric conditions (curve, slope, etc.), and traffic conditions (volume, speed, density, etc.). The major difficulty of this type of study was to assign the

VMS randomly, which was nearly impossible in practice. This was because, virtually all the VMS were installed “on purpose” to solve a specific traffic problem in a particular road section (Haghani et al., 2013). “On-and-off” study focused on the comparison of the crash probabilities when the VMS was active and inactive. This type of study was typically single-VMS based, and the primary idea was to seek the relationship between the deployment of VMS and crash occurrence, with other possible causal factors under controlled (Haghani et al., 2013). The study described in this paper follows the concept of “on-and-off” study.

“Before-and-after” study compared the crash rates in the specific road segments before and after the installation of a VMS. This type of study was generally recognized to be the “naive” analysis, which seldom considered possible causal factors other than the presence of VMS (Mounce, 2007). Using a simulation model was another way to assess the safety effectiveness of VMS (Hoye, 2011). However, compared to the other types of study, simulation method was less persuasive unless the models were calibrated for the specific study areas.

Difficulties do exist when performing the safety evaluation of VMS. Previous studies tended to underestimate the difficulties and used inaccurate or incomplete information when analyzing the impact of VMS on road safety. First, the definitions of VMS impact regions were not consistent. VMS impact regions are the areas that road users could be potentially affected by the deployment of sign messages. The definition of the region is primarily important for crash data collection, which is significantly associated with the research results. However, there is no consistent definition according to the existing publications. Second, the crash data used in most existing studies were obtained from police reports or similar systems operated by local departments of transportation. These often have approximate estimates of crash location and time. The locations and times of the crashes are critical in safety evaluation to determine if they occurred within the VMS impact regions and when the VMS were active. The direct use of the “second hand” crash information could result in a bias of the results. Third, other crash causal variables were not considered. The deployment of VMS probably is not the only causal factor to the crashes. Other potential risk variables, such as traffic conditions, road curvatures, weather

conditions, need to be recognized and controlled in the analyses. Earlier studies, especially “before-and-after” studies, seldom included a comprehensive collection of control variables in analysis models. On the other hand, few previous studies could draw a comprehensive conclusion that the deployment of VMS, rather than the downstream shockwaves, was the cause of the significant speed change. Short-term fluctuations in traffic are considered dangerous to upstream traffic as well, and the VMS is not supposed to result in the fluctuation. In short, it is necessary to conduct a more rigorous safety evaluation with the difficulties stated above overcome.

### **1.3 Safety Evaluation Methodologies in this Study**

This study focused on the safety evaluation of VMS. It was conducted by evaluating five VMS along Interstate-94 from Minneapolis to St. Paul. The evaluations were performed to test two hypotheses. The first was that vehicle speeds in the VMS impact regions showed no difference when a VMS was active or inactive. For example, drivers did not slow down to obtain more time reading VMS messages. The second hypothesis was that, the probability of crashes in the VMS impact regions was not associated with the deployment of VMS. In other words, the VMS distraction was not a causal factor to crashes. The two hypotheses were tested separately in Chapter 4 and Chapter 5.

The analysis for the VMS impact on traffic speed was conducted using normal linear models. The primary idea was to test the average change of speed affected by the deployment of VMS, and various possible causal factors, such as messages displaying time and weather conditions. In order to eliminate the noise of speed change caused by the events described in the VMS messages, cases potentially influenced by downstream shockwaves were identified, and removed from the analysis. In addition, a difference scores model and an analysis of covariance model were compared to investigate the interactions among the initial speed that vehicles arrived at the VMS impact regions, and the other explanatory variables.

A case-control study was performed to analyze the association between the deployment of VMS and crash occurrence. The unadjusted and adjusted odds ratios were estimated respectively with the 2×2 contingency tables and logistic regression models. The logistic models included numerous control variables, such as traffic metrics, weather

conditions, and crash diagrams. The impact of VMS was investigated by comparing the probabilities of crashes during the VMS were active and were blank, under the condition that the other possible causal factors were controlled. In addition, rather than using the aggregated data, the models estimated the odds ratios for different types of VMS messages classified in terms of acknowledged criteria.

#### **1.4 Organization of the Thesis**

This thesis has six chapters in total. Chapter 2 describes the previous publications on the evaluation of VMS, crash-prone traffic studies, and the impact of adverse weather on crashes. Chapter 3 describes the study scope, data collection, and data preparation. Chapter 4 analyzes the influence of VMS deployment on the change of vehicle speed in their impact regions. Chapter 5 studies the association between the deployment of different types of VMS messages and the probabilities of crash occurrence. Chapter 6 summarizes the findings from the studies, and highlights the achievements and limitations.

## **Chapter 2: Literature Review**

The literature review stated in this chapter includes three parts. In the first part, the previous studies on the evaluations of VMS are summarized. Those evaluations were separated into qualitative evaluation, and quantitative evaluation. The safety evaluation generally belonged to the latter. Since the impact of VMS on road safety is not recommended to study without considering the effect of traffic conditions, a review of studies on crash-prone traffic conditions was elaborated in the second part. In addition, in order to identify the impact of VMS during different types of weather conditions, a review of adverse weather influence on crashes was discussed in the third part.

### **2.1 Evaluations of Variable Message Signs**

#### **2.1.1 Qualitative Evaluation**

Qualitative evaluation of VMS is also named as driver survey oriented evaluation. Driver acceptance and comprehension are important as parts of the performance of VMS, and driver survey is one of the proper ways to address the issue. For this kind of evaluation, researchers mostly considered the percentage of drivers who were satisfied with VMS messages, how drivers trusted the messages, whether they reacted after reading the messages, and their expectations to VMS performance. The survey results could provide researchers and decision makers a general view of how VMS affect drivers' behavior. The types of survey can be summarized as general survey, stated preference (SP) survey, and professional survey, which is also called "focus group discussion" according to Mounce (2007).

##### *a. General Survey*

General survey requires drivers to recall their sense of satisfaction and reactions to VMS messages based on their driving experience. For example, drivers are requested to complete a take-home questionnaire that assesses their familiarity to VMS and their reactions to specific messages. Normally, this kind of survey is questionnaire-based by mail, by phone, in field, online or in person. The Intelligent Transportation Systems (ITS) Program at the University of Wisconsin (2004) did a driver survey on evaluating VMS in

Wisconsin in 2001. The survey sheets were randomly sent to 22 counties in Wisconsin where VMS were widely used on freeways and trunk highways, and 218 sheets were mailed back for study. Although the survey was limited to one state of United States, it helped researchers come to some interesting conclusions, such as 15% of drivers never heard of VMS. Tay (2008) did a similar driver survey in Canada. The survey was given to two sample groups: one group mostly consisted of students at campus who were supposed to have limited driving experience; the other mostly consisted of skillful drivers. The combined results showed that the majority of drivers (86%) had looked at the VMS, and 70% of them did think about the messages after reading. A general survey is mainly used to identify drivers' perceptions towards the VMS in a qualitative way. The objective of the survey is to give researchers and decision makers a general view of drivers' possible reactions to the VMS messages. Typically, general survey is costly and time consuming, but it can capture valuable information on driver characteristics compared with the other evaluation methods. Some studies (Wareman, 1997) also used survey data to analyze the impact of VMS on traffic characteristics such as diversion rate and traffic flow pattern.

*b. Stated Preference Survey*

Stated preference (SP) survey requires participants to respond to VMS messages in specific, well-defined hypothetical situations. This technique is widely used in economic and marketing fields, and has become popular in traffic (Mounce, 2007). It takes less effort than a general survey but emphasizes specific research topics. For example, Wareman (1997) used a stated preference approach to study the drivers' route choice influenced by VMS. During the survey, participants (drivers) were showed screen-pictures of the freeway and VMS ahead, and their responses of route choice were recorded. Survey results showed that messages of "Qualitative descriptions of delays" resulted in higher probability of route choice than the other vague delay messages; and visible queues was also a major cause of route choice that might not be attributed to the VMS messages. Chen et al. (2008) did a similar stated preference investigation on drivers' reactions to VMS route guidance messages. The main findings were that only 7.02% of drivers changed their current routes under common congestion (21.45% under serious



congestion), and experienced drivers always preferred their former routes to rerouting. The main advantage of an SP survey is it can capture useful information of drivers' reactions under some specific traffic conditions, which offers a better option to analyze drivers' behavior. However, the validity of respondents' choices is open to question. Drivers' route choices vary from different time ranges (weekday peak hours or weekends) and the familiarity with the neighboring road system. The reactions towards simulated situations, such as "screen traffic condition", may not represent drivers' willingness to choose routes in real traffic conditions. Adjustment factors must be derived and applied for the further analysis if SP survey is performed. Using the survey results directly may result in bias in conclusions.

*c. Professional Survey*

Professional survey, also named as "focus group discussion" according to Mounce (2007), is another useful survey method gathering qualitative information from specific groups of people. Focus group discussion is broadly used in social science research, and it is the combination of focused interview and group discussion (Kallbekken and Marianne, 2010). The latter is generally applied in traffic studies. Tay (2008) organized a professional group discussion before the general survey. Participants were all professionals in traffic and transportation fields. The topics were focused on the awareness of VMS messages, different uses of VMS, and display of road safety messages. Participants in the discussion presented their opinions and comments in professional ways, and all their points of views were recorded and organized. The results of group discussion were then used to guide the general driver survey. Focus group discussion can explore opinions from both professional people and the public. Moreover, it is beneficial to the further qualitative studies by helping researchers make assumptions, and identify the test variables (Kallbekken and Marianne, 2010).

The driver survey-oriented evaluations are important, and many studies evaluating the performance of VMS referred to survey data (Hoye, 2011; Chen, 2008; Mounce, 2007). With survey data, Lee (2007) used a fuzzy aggregation method to estimate the satisfaction of drivers toward the delay and travel time information provided by VMS. Warenman (1997) analyzed the drivers' route preference corresponding to VMS

messages by constructing a multinomial logit (MNL) model based on survey data.

### **2.1.2 Quantitative Evaluation**

#### *a. Diversion Rates Evaluation*

One of major functions of VMS is to guide drivers to alternative routes when certain non-recurring congestion (caused by crash, disabled vehicles, etc.), and recurring congestion (caused by daily normal congestion) happens. The effectiveness of VMS on route guidance can be tested by observing the change of diversion rates before and after the display of VMS messages. The approaches to diversion rates evaluation consist of statistical analysis and simulation models.

For statistical analysis based study, traffic data were usually collected from driver surveys, loop detectors, video cameras, etc. Survey data sometimes were utilized in certain studies (Hoye, 2011) to estimate change of diversion rates caused by VMS, but only in a qualitative way. Loop detector data and video camera data were widely used in diversion rates evaluation. Schroeder and Demetsky (2011) fitted a generalized linear model estimating the diversion rates based on archived traffic loop data and categorical VMS message data on I-95 to I-295 in Virginia, from 2005 to 2008. The study was performed with highly disaggregated time, traffic, and VMS data. The study found that higher diversion rates were detected when specific diversion route information were contained in the VMS messages, especially in off-peak hours. With the empirical loop detector data, Levinson and Huo (2003) constructed a statistical model showing that VMS had impact on the diversion rates for 10 minutes after messages were activated. Loop detector data have the advantage that they reflect the aggregate pattern of traffic characteristics, but not for individual drivers. However, video camera data or Bluetooth data could capture the characteristics of individual vehicles like route choice and accurate travel speed. A study conducted by the University of Maryland (Haghani et al., 2013) using Bluetooth sensor technique showed that diversion rates increased 5 to 20 percent when alternative route guidance was displayed on the VMS panels.

Instead of using real data evaluating the impact of VMS on route choice, the application of simulation models is also a common approach. The power of simulation is that researchers can control parameters and repeat experiments multiple times (Kolisetty

et al, 2005), and optimal results can be obtained from the iterations. However, those results may or may not be reasonable and feasible in reality, and validation efforts are always necessary. Chang et al. (2002) developed a simulation model with Traffic Simulator to estimate the optimal detour rate. The goal of the study was to minimize the delay caused by the non-recurrent congestion on freeways by maximizing the use of VMS. Simulation results revealed that reduced roadway capacity, incident clearance time, and traffic volume were key factors that could affect the detour rate. Chen et al. (2008) also conducted a simulation model using VISSIM, which is a multi-modal traffic flow simulation software, to evaluate the effectiveness of VMS on route guidance and congestion. Shang and Lu (2009) studied VMS route guidance using cell transmission model (CTM) simulation. The study revealed that VMS route guidance was more effective if the traffic demands were higher. The studies stated above simulated driver behavior using multiple software and models in order to evaluate the impact of VMS on diversion rates. However, constructing a simulated situation to measure driver behavior is another approach. Harder, et al. (2003) conducted a study assessing the diversion rates affect by a specific VMS message with a fully-interactive, PC-based STISIM driving simulator. The results showed that around 56% of participants chose to divert after seeing a specific “Exit” Message. However, simulation-based studies can only be performed on specific research topics and should not be simply compared without identifying study boundaries.

*b. Travel Time and Delay Impact Evaluation*

Travel time and delay-based evaluation is one of the most important parts for the mobility evaluation of VMS. Texas Department of Transportation believed that the reduction of delay and risk caused by non-recurring incidents was a major objective of VMS (Mounce, 2007). Similar to the evaluation of diversion rates, the methodologies for travel time and delay-based evaluation are typically divided into two aspects: the application of statistical models and simulation models. The former methodology generally depended on freeway loop detector data. Individual vehicle analysis is neither straightforward nor efficient to process when the study objects are travel time and delay on a long segments or a whole corridor. Levinson and Huo (2003) utilized loop detector

data to conduct a before-and-after evaluation of VMS on freeway travel time and delay reduction. In terms of the statistical results, total delay decreased after the installation of VMS, but the effect on total travel time was not obvious.

Simulation models for evaluating VMS effects on travel time and delay reduction have been widely used recently. As stated above, repeatable process and optimal outcomes are two major advantages of simulation models. A study in Trondheim, Norway (Hoye, 2011) estimated the potential impact of VMS messages on travel time reduction based on CONTRAM simulation model. Results indicated that travel time was likely to reduce but was associated with increase in crash rates. A Malaysia study (Roshandeh, Puan, 2009) also took advantage of simulation models to estimate the traffic characteristics (travel time, delay, occupancy, and gap) corresponding to the installation of VMS. In a before-and-after comparison, travel time went down after the VMS installation, which was consistent with Hoye's study (2011). Shang and Lu (2009) using a Cell Transmission Model (CTM), also confirmed that total travel time decreased with VMS route guidance. Although simulation models are powerful, researchers have to be cautious when choosing models and making assumptions. Some assumptions in the simulation models might not reflect reality. For example, a Norwegian study (2011) assumed that all the drivers were familiar with the road system in Trondheim; Shang's study (2009) assumed that drivers could only choose an alternative route rather than other options like postponing or canceling the trip, or using other transportation means. Some assumptions are probably beneficial to the interpretation of the results, but are not practical. Moreover, calibration and validation are needed for all the simulation models.

In summary, both the application of statistical models and simulation models can assist the evaluation of the VMS on corridor travel time saving and delay reduction. Although each method has its merits and demerits, most of the results confirmed that the deployment of VMS has a positive impact on the reduction of corridor travel time. However, argument still exists on whether it is beneficial for delay reduction or prevention.

### *c. Speed Impact Evaluation*

The evaluation of VMS impact on short-time speed fluctuation is commonly considered as an indirect way to evaluate its effect on freeway congestion and accidents. Generally, the methodology for speed evaluation is “on-and-off” analysis with the application of statistical models. Based on Bluetooth sensor technology, Haghani (2013) performed a study on short-time speed variation by different types of messages displayed on the VMS panels. Speed data were collected for five minutes at one-minute intervals before and after the display of each VMS message, and paired t-tests were utilized to test the significance of speed change before and after the VMS deployment. Results revealed that 17% of the total cases showed a significant decrease of speed with an average of 3.0 mph when a VMS was activated, and during the “danger/warning” message period, speed was generally lower than “blank” message period. Erke et al. (2007) conducted a field test and video observation study on speed variation caused by display of VMS messages. In this study, speed reduction and braking maneuvers were observed by looking at the rear-lights of individual vehicles. By controlling the messages type of VMS, they found that the average speed reduction was around 4.0 mph, and the overload of VMS information could result in the reduction of speed as well. Moreover, chain reactions of braking were investigated, and results confirmed that following vehicles were likely to either brake or change lanes to avoid collisions if the leading car reduced speed due to the impact of VMS. Harder et al. (2003) conducted a study that investigated the driver behavior under the impact of some specific VMS messages with a PC-based STISIM driving simulator. Results revealed that more than 20 percent of participants slowed down no less than 2 mph.

In short, according to the previous studies, speed decreased when certain messages were activated (Haghani et al., 2013; Erke et al., 2007; Harder et al., 2003). This indicates that VMS could cause short-term traffic fluctuation, and the shockwaves caused by the speed change (if great enough, for example, 10 mph) could lead to an increased potential for crashes. However, the entire mechanism affected by VMS is complicated, and there is lack of researches providing strong interpretations on that.

#### *d. Safety Impact Evaluation*

As stated in the introduction, the types of safety evaluation of VMS consist of “with-and-without” study, “on-and-off” study, “before-and-after” study, and simulation based study. In some studies, those types of evaluations were used together. An example of “with-and-without” study conducted by Haghani et al. (2013) randomly picked 70 segment samples on I-95 in Maryland. A portion of the sample segments had VMS deployed. Then, Poisson and negative binomial regressions with the dependent variable being crash counts and independent variables of AADT, existence of interchanges, and impact area were used to analyze the data. An unbalanced two-way ANOVA showed that AADT and interchange were important factors for accidents, but not presence of VMS. Then an “On-and-Off” study, which analyzed the probability of crash occurrence with or without messages showed on VMS, was conducted. Fifteen VMS samples were selected and a one-way ANOVA analysis was performed. Results revealed that crash rates were lower when VMS is active compared to no messages displaying. “Before-and-After” studies were widely used in safety evaluation of VMS. However, mostly this type of evaluation was applied to practical projects rather than research projects. One example is “Guidance of Evaluation of VMS” in Texas (Mounce, 2007), which developed a reduction ratio to normalize the comparison of effects. Based on the reduction ratio, the function for estimating the change of crash ratio before and after the installation of VMS was constructed. Other than traditional approaches to assess the localized safety evaluation of VMS, some studies also utilize simulation models for crash rates evaluation (Hoye, 2011). However, simulation models were not popular on safety evaluation of VMS due to the difficulty in predicting crashes and incidents.

In summary, the installation of VMS is not supposed to increase the potential risk of crashes. Some studies confirmed the contribution of VMS on preventing crashes, but some did not. Due to the concern of accuracy and difficulty of safety evaluation of VMS stated in Chapter 1, there is no consistent conclusion on the association between the deployment of VMS and crash occurrence.

## **2.2 Studies on the Crash-Prone Traffic Conditions**

Previous studies on the crash-prone traffic conditions indicated that, the fluctuation in

traffic flow, speed, and density adjacent to the crashes had impacts on the probability of crashes. This implies that, the deployment of VMS is probably not a single risk factor when evaluating its performance on freeway safety. Researchers are required to have a deep understanding of the mechanism of crash occurrence, and to apply the findings to VMS evaluation. Currently, the studies of crash-prone traffic conditions were mostly real-time traffic data based, and mainly focused on the crashes potentially caused by upstream or downstream short-time variation of speed, flows, and density.

### **2.2.1 Traffic Speed and Crashes**

Upstream and downstream speed variation prior to crashes is commonly regarded as a risk factor. This fact has been confirmed by many researchers. Oh et al. (2001) conducted a study estimating the crash likelihood utilizing a nonparametric Bayesian model. The traffic conditions 30 minutes before the crashes were split into two periods: normal conditions and disruptive conditions. The standard deviation of 5-min average speed was chosen as the only accident indicator based on a t-test analysis. Model estimation results revealed that more crashes were associated with a higher standard deviation of speed upstream. Another study by Hourdos et al. (2007) built a crash likelihood model based on the real-time crash data extracted from video cameras, and then numerous traffic metrics were calculated. A generalized linear regression for binary response was applied in the model and results showed that speed variation did affect the crash likelihood. Moreover, Lee et al. (2002) developed a crash prediction model using a log-linear analysis. In addition to real-time traffic data, weather, geometric, and time of day were controlled in the model. One of the main findings was that variation of speed difference across lanes was a statistically significant predictor for estimating crash frequency. Zheng et al. (2010) also studied the impact of traffic oscillations on freeway crash occurrence with a matched case-control design that utilized high-resolution traffic and crash data. Based on the case-control samples, conditional logistic regression models were developed and results revealed that standard deviation of speed was more significant than average traffic states, especially in congestion conditions.

Instead of using standard deviation of speed as the indicator detecting crash-prone conditions, coefficient of variation in speed (standard deviation speed/average speed) is

also an applicable index. One of the representative studies was conducted by Abdel-Aty and Pande (2004) with the methodology of Bayesian classifier and probabilistic neural network (PNN). Based on classified historical crash data and traffic data, outcomes from difference analysis strategies revealed that logarithms of the coefficient of variation in speed stood out to be the most significant predictor.

Since speed variation prior to the crashes is potentially associated with crash frequency on freeways, two ways can be considered to assess the safety evaluation of VMS. One is to construct a statistical model with crashes as response, and independent variables that include both factors of VMS and speed variation prior to crashes. By controlling the traffic conditions, the impact of VMS can be interpreted by the model results. The other approach is to study the impact of VMS on short-time traffic fluctuation by analyzing the upstream and downstream speed variation corresponding to the messages displayed by VMS. For the second approach, an assumption is that drivers do not slow down for reading any types of the messages. Haghani (2013), Erke et al. (2007), and Harder et al. (2003) tested this hypothesis, and confirmed that VMS had an impact on driver behavior that could result in speed reduction. However, they did not come to a comprehensive conclusion that whether a simulation result could represent the actual driver behavior, whether the speed variation was indeed caused by VMS, and whether this was a potential risk to road safety.

### **2.2.2 Congestion Levels and Crashes**

Other than speed variation prior to the crashes, congestion level is also correlated with the probability of crashes on freeways. According to the previous studies, there was a positive correlation between crash rates and traffic volume. In the study of Yeo, et al. (2012), which analyzed relationships between different traffic metrics and crash occurrence, four section-based traffic states (free flow, back of queue, bottleneck front, and congestion) were defined, and a formula of crash involvement rate was developed. Based on three-year period crash and detector data, the rate was projected to each traffic states by 5 mph speed interval, and the result indicated that the probability of crash was around 5 times lower in free flow condition than the other conditions. On the other hand, crash severity is also a critical index on the evaluation of road safety. Some research



results revealed that crash rates increase when traffic volume was high, but the severity of crashes decreased. Quddus (2009) studied the association between the level of congestion and severity of crash with disaggregated crash records. Ordered logit models, heterogeneous choice models, and generalized ordered logit (partially constrained) models were used in the analysis to test the significance of each independent variable. The key finding was that crash severity decreased with increasing traffic flow, but not level of congestion.

Since the probability of crash possibly depends on traffic flows, the measure of congestion levels should be considered when evaluating the safety performance of VMS. Traffic patterns during the peak and off-peak hours, and weekday and weekends are not the same. Typically, crashes are more probable in congested conditions during weekdays and in free flow conditions during weekends (Yu and Abdel-Aty, 2013). Thus, traffic flow adjacent to the crashes needs to be controlled in the localized evaluation of VMS.

### **2.2.3 Density (Occupancy) and Crashes**

Traffic density (or occupancy) is also detected to be one of the significant crash predictors. In a crash-prone traffic conditions study conducted by Shively et al. (2009), a semi-parametric Poisson-gamma model was utilized, and the model results suggested that the amount and density of traffic were key factors that influenced the variability of crash rates. In addition, Abdel-Aty et al. (2004) proved that 5-min average occupancy observed upstream of a crash had significant impact on crash occurrence, by developing a crash-likelihood prediction model, with real-time traffic flow variables. Moreover, the study of Lee et al. (2002), which utilized a log-linear model, also confirmed that traffic density and variation of speed were both statistically significant for crash frequency prediction. Since density is not easy to be measured directly, lane occupancy is used as a substitute, which can be accessed by loop detectors. Either of them should be considered in the evaluation models.

In summary, as reported by previous studies, the variation of three principal macroscopic parameters (volume, speed, density) that characterized the traffic stream, all have impacts on the probability of crashes. Hence, the “naive” conclusions that stated the association between crash rates (or crash probability) and VMS deployment without

considering the variation of traffic metrics were not defensible. The crashes occurred in the VMS impact regions can be attributed to the distraction of drivers affected by VMS messages. However, it could also result from the fluctuation of traffic, such as large downstream speed variation, but not the VMS. In addition, some crashes were possibly caused by both. On the other hand, if the VMS messages potentially make drivers slow down for reading, the VMS has the power to produce shockwaves that could lead to further upstream crashes, which is also complex to interpret. Therefore, the definition of VMS impact region for each study is also important in order to avoid linking crashes to impact factors incorrectly. Anyway, analyzing traffic conditions prior to the crashes is indispensable for safety evaluation of VMS.

### **2.3 Impact of Adverse Weather on Crashes**

Weather conditions are commonly used when developing crash-likelihood prediction models. According to Federal Highway Administration (FHWA), weather-related crashes, which are crashes that happened in the condition of the rain, sleet, snow, fog, wet pavement, snowy or slushy pavement, or icy pavement, are 24 percent of total crashes (Pisano et al., 2008). Adverse weather conditions like snowfall and rainfall lead to the reduction of the friction factor for pavements and visibility of drivers, which could directly increase the probability of crashes. However, it is not a simple chain effect that adverse weather is associated with more crashes. Researchers found that crash frequency tended to decrease in adverse weather conditions, possibly because drivers were likely to lower the speed and to raise vigilance (Strong, 2010; Eisenberg, 2005). For VMS safety evaluation, the involvement of weather factors can affect whether VMS has different impacts on traffic flows and crash frequency in adverse conditions than normal conditions. Moreover, the control of weather factors is also beneficial to study the association of crashes and the impact of VMS.

The impact of snow on crash occurrence has been widely studied since 1970s, and its negative influence on traffic safety has been proved by many research results. A study in Canada (Mende, 1982) also found that crash rates could increase from 30 to 140 percent in significant snowfall conditions. A similar study addressed by Andreescu (1998) revealed that number of crashes increased sharply with the increase of snowfall. However,

the study of Eisenberg et al. (2005) indicated that, although snowy days had more nonfatal-injury and property-damage-only crashes, the fatal crashes were less than dry days.

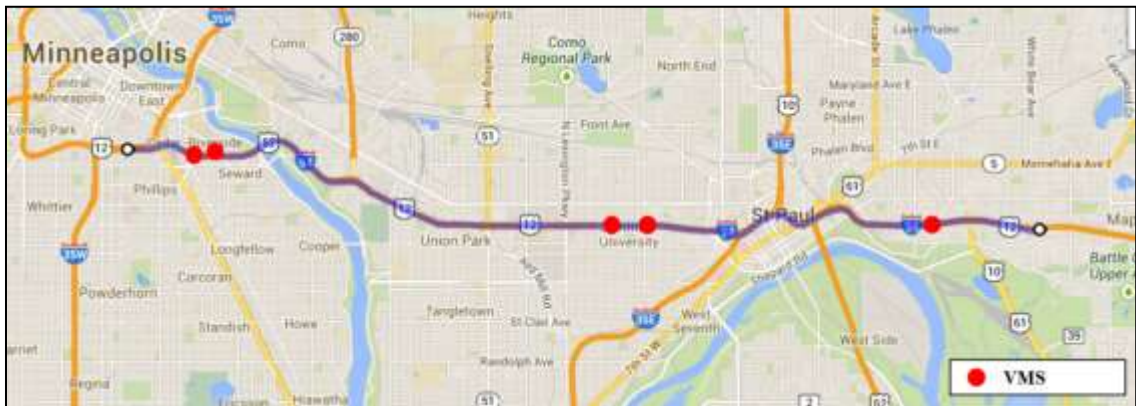
Rainfall is also a commonly proved risk factor to crash occurrence. Abdel-Aty and Pemmanaboina (2006) found that a rain index was a crucial predictor for the crashes by performing a matched case-control logit model with traffic loop data and rain data. A study conducted by Yu et al. (2013) indicated that weather condition variables, especially precipitation, were statistically significant in crash occurrence models. Furthermore, Xu et al. (2013) found that rainfall interacted with speed difference between upstream and downstream stations on the crash prediction results.

Other weather impact factors such as fog, snow, and frozen pavement are not as commonly studied as rain and snow. However, this does not mean that their influence on crash occurrence is not significant. A study conducted by Abdel-Aty et al. (2011) using a four-year period crash and weather data from Florida found that multi-vehicle involved crashes with more severe injuries were likely to happen in foggy or smoky weather conditions. In additions, these crashes are usually head-on and rear-end crashes. In short, safety evaluations of VMS should be conducted under various weather conditions.

## Chapter 3: Data Collection and Preparation

### 3.1 Definition of Study Scope

The corridor selected in this study was Interstate-94 from downtown Minneapolis to downtown St. Paul in Minnesota. Five VMS sites were used for analysis (three on the westbound, and two on the eastbound, see FIGURE 1), and the corresponding crash data, crash-prone traffic data, and weather data were collected. The study period was ranging from January 1, 2006 to December 31, 2012.



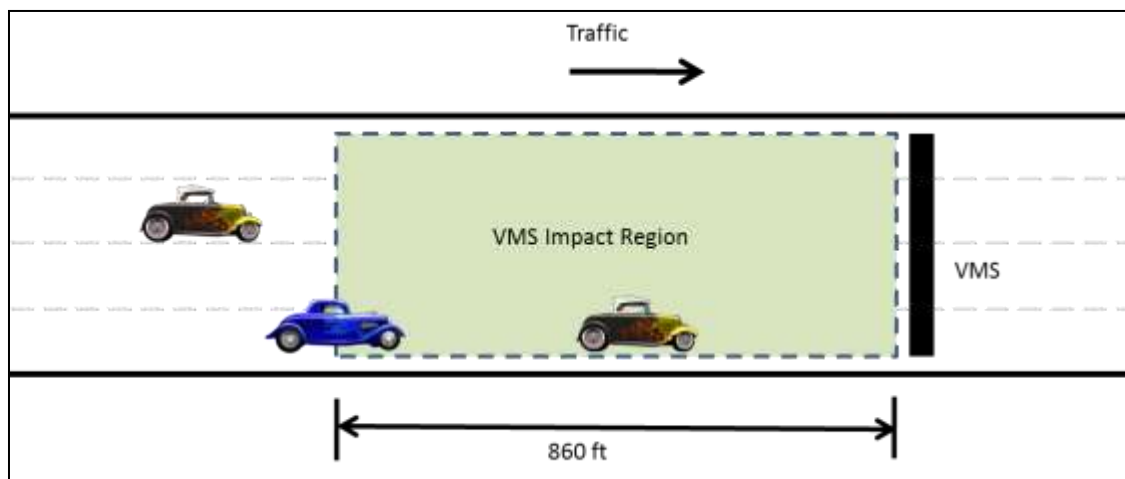
**FIGURE 1 Study Corridor and VMS Locations (Screenshot from Google Maps)**

The VMS impact region is defined as road segment starting from a VMS panel to 860 feet upstream away. (FIGURE 2) The reason for choosing this space interval was based on a Manual on Uniform Traffic Control Devices (MUTCD) requirement that the signs should be legible to drivers at a minimum of 860 feet (Harder et al., 2003). That is, all drivers in the impact region should have notified the VMS, although some may have responded earlier. The road segments for the five defined VMS impact regions had similar characters, and there were no obvious horizontal and vertical curves applied to those segments. For the accuracy of the research, crash data were collected only in those regions. Then, loop detectors were selected based on the VMS impact regions and locations of crashes. The distance between the crashes and detectors (both upstream and downstream) was within 0.5 mile, in order to guarantee that the impact of crash on traffic

could be captured. Weather data were collected from weather stations within 5 miles from the VMS impact regions.

### 3.2 VMS Log Data Collection

VMS logs are records of the set of messages displayed by VMS panels. Typically, each message consists of three items of information: a problem statement (crash, incident, debris, etc.), a location statement, and an effect statement (lane closed, expecting delay, etc.) These messages are organized and saved in the VMS logs along with an event ID, an event date, a description, a device ID, the message, and the roadway, road direction, cross street and operator. All the VMS logs used in this study were requested from MnDOT.



**FIGURE 2 VMS Impact Region**

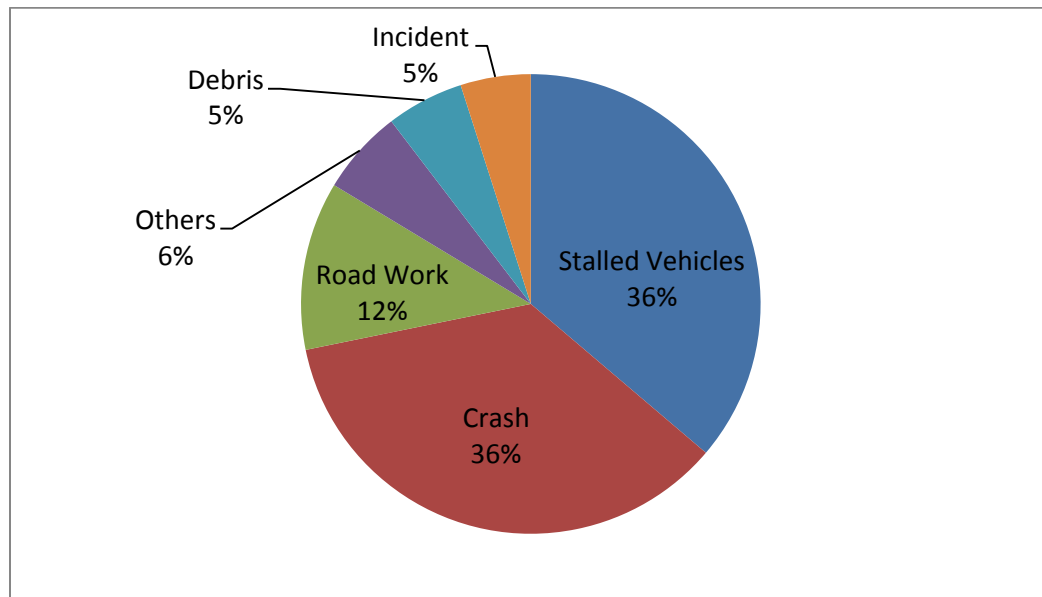
In this study, five VMS devices were used along Interstate-94 between downtown Minneapolis and downtown St. Paul. Seven-year period log data were collected (January 1, 2006 to December 31, 2012). The detailed information of VMS devices used in this study is summarized in TABLE 1.

For research purpose, the major information needed for each log record were the starting and ending time of each message, and the detailed message description. VMS logs do not have the ending time of each message, but have a column called “description” that tells the status of VMS. The unique marks in this column are “Sign DEPLOYED” or “Sign CLEARED”, and the latter was used for deriving the ending time of each message.

With the starting time and the ending time, the durations of the messages were calculated. However, some of the “Sign CLEARED” marks were missing or not reasonable in the logs, for example, it took 7 days to clear a crash. These records were detected, and removed.

**TABLE 1 Detailed Information for Selected VMS**

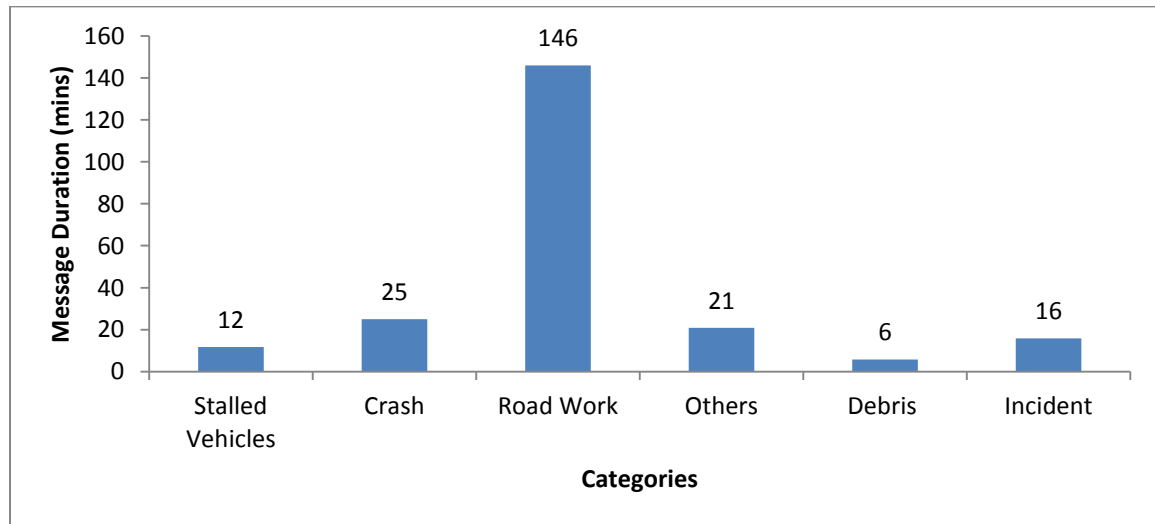
Device ID	Corridor	Direction	Cross street	True Miles
V94E08	I-94	Eastbound	25th Ave	235.296
V94E11	I-94	Eastbound	Victoria St	241.238
V94W05	I-94	Westbound	Earl St	244.948
V94W07	I-94	Westbound	Dale St	240.969
V94W09	I-94	Westbound	25th Ave	235.142



**FIGURE 3 The Frequencies for Each VMS Message Type**

Based on the VMS logs for five selected devices, messages in the study regions could be summarized into five categories, which were “Crash”, “Incident”, “Stalled Vehicle”, “Debris on Road”, “Road Work”, and “Others”. The frequency of each category is showed in FIGURE 3, and median duration of each category is displayed in FIGURE 4. According to the figures, messages that display “Crash” and “Stalled Vehicles”

contributed most of messages counts (72% of total messages counts). The median clearance time for each category estimated by VMS logs was within 30 minutes except “Road Work”. This finding is reasonable because the response of the freeway incidents is usually immediate and effective, especially on the high-occupied road segment. However, the duration of roadwork is not reliable, and depends on different workloads. Its median durations are not as representative as the others are.



**FIGURE 4 The Median Durations for Each VMS Message Type**

**TABLE 2 Haghani's Grouping Table for VMS Message Types**

Type	Examples of Displayed Messages
Blank	No message
Danger/Warning	Incidents, Disabled Vehicles, Non-recurring Slow-Downs, Roadway Debris, Unplanned Lane/Tunnel/Bridge Closures
Informative/Common	Roadwork Closures, Major & Minor Delays, Congestion, Travel Time, Other travel-related messages (Fog, Ice, Snow Plowing, Major Events)
Regulatory/Non-Traffic	Work Zone Speeds, Seatbelt Use, Cell Phone Regulations, Motorcycle Awareness, Amber Alerts, Homeland Security Messages

Although there were five major categories for messages displayed on the VMS panels, some categories in different groups could have similar impact on traffic. According to the previous studies, various ways were used to group the VMS messages based on the level of impact, duration, frequency, etc. In this study, messages were grouped into three categories, which aligned with the classification methodology provided by Haghani (2013), whose grouping rules are displayed in TABLE 2.

In summary, seven years of log data were collected (January 1, 2006 to December 31, 2012) for five selected VMS devices, and 7849 messages were available for analysis.

### **3.3 Crash Data Collection**

#### **3.3.1 Crash Data Sources**

Crash databases available in this study were from three data sources. The first one was the Minnesota Department of Transportation Incident Logs/Computer Aided Dispatch (CAD) system, the second one was the Minnesota Department of Public Safety (DPS) Crash Database, and the third one was the VMS messages logs.

The major sources used for research purpose usually are the first two databases. The MnDOT Incident Logs were transitioned to the State Patrol CAD system for collecting freeway incidents and incident clearance times since August 2008 (FHWA, 2009). The CAD system records freeway crashes from police patrols and the Regional Transportation Management Center (RTMC) office. The DPS crash database is a summary of police reports, which are prepared by Minnesota State Patrol and local law enforcement departments (FHWA, 2009). Compared to the CAD data, DPS data has more detailed information on the nature of crash. The VMS database is actually the set of messages displayed on the VMS panels, which describe the information on crashes. It is essential and effective when studying the traffic impact for each incident. The major information that each database provides is summarized in TABLE 3.

#### **3.3.2 Advantages and Limitations**

One of the major advantages of the CAD system is that it provides the crash clearance time and ending time. Moreover, the information is real-time, and can be used for



research immediately. However, the location information reported in the CAD system is not always accurate. In addition, it does not contain the detailed information for each crash (FHWA, 2009), and only freeway crashes are documented. In comparison with CAD data, DPS data contains adequate information for each crash, such as severity, type, road surface and even weather conditions. Furthermore, crashes in DPS data are recorded 24 hours a day, 7 days a week, and 365 days a year (FHWA, 2009). Nonetheless, the main limitation of DPS database is it only contains the records of crashes that were reported. For example, some minor crashes or crashes without the mediation of police officers are not included in the DPS system. Apparently, CAD system has a higher level of representation of freeway crash than DPS database, but DPS data are more accurate in details.

**TABLE 3 Detailed Information Provided by Different Crash Data Sources**

Information	CAD System	DPS Database	VMS Database
Incident Type	Yes	Yes	Yes
Accurate Location	Sometimes	Sometimes	No
Starting Time	Yes	Yes	Yes
Ending Time	Yes	No	Sometimes
Incident Clearance Time	Yes	No	No
Lane Blocking	Yes	No	Sometimes
Severity Of Crash	No	Yes	No
Crash Type	No	Yes	Sometimes
Crash Diagram	No	Yes	No
Road Surface	No	Yes	No

For crash information extracted from the VMS logs, its accuracy is lower than the other two data sources. Nonetheless, the objective of displaying VMS messages is to provide information of traffic conditions and road guidance to drivers, rather than crash data collection. However, crash messages in the VMS logs contain information of the traffic impact caused by a single crash. For example, if a crash was cleared but it still had impact on traffic, VMS log would display the message until traffic was back to normal.

Thus, in this study, although VMS crash data were not utilized directly for analysis, it was valuable to validate the information provided by the other two crash sources when the crash information was missing or obscure.

### 3.3.3 Crash data Used in this Study

DPS data were primarily used for preparing the crash database in this study. The data could be accessed from Minnesota Crash Mapping Analysis Tool (MnCMAT) with an authorized license. Based on the location of the each VMS site and the defined VMS impact region, crash records could be filtered by their “Mile Mark”, and be extracted directly from the databases. For each VMS site, seven-year period crash data from January 1, 2006 to December 31, 2012 were identified. In total, 183 qualified crashes were collected in five VMS impact regions. (See TABLE 4)

**TABLE 4 Number of Crashes in Each VMS Impact Region**

Region	VMS ID	Number of Crashes
1	V95W05	28
2	V94W07	51
3	V94W09	59
4	V94E08	35
5	V94E11	10
Total		183

**TABLE 5 Crash Diagrams and Codes**

Code	Description	Counts	Code	Description	Counts
0	Not specified/null value	0	7	Ran off road - right side	7
1	Rear end	115	8	Head on	1
2	Sideswipe - same direction	34	9	Sideswipe - opposing	0
3	Left turn	0	90	Other	16
4	Ran off road - left side	8	98	Not applicable	0
5	Right angle	2	99	Unknown	0
6	Right turn	0			

The critical attributes used in this study included crash identification number, reported time, crash mileposts, and diagrams. Crash diagrams are useful information when predicting crash, or identifying crash causal factors. (Christoforou et al., 2011) The analysis with aggregated crash types may cause the bias of the results. According to the MnCMAT database, the crash diagrams are coded in TABLE 5. A distribution of crash counts used in this study was followed the description of crash diagram. Rear-end crash has the highest frequency (63%).

### **3.4 Traffic Data Collection**

Since many studies (Oh et al., 2001; Hourdos et al., 2007; Abdel-Aty and Pande, 2004; Shively et al., 2009) indicated that the probability of crash was higher at the road segments where short-time fluctuations in downstream or upstream traffic was strong, it is important to consider the traffic conditions prior to the crashes when evaluating the impact of VMS. In this study, the causal relationship of VMS and crash occurrence was studied with the traffic metrics being controlled.

#### **3.4.1 Crash Time Estimation**

Extracting information of traffic conditions prior to the crashes is mainly based on the time when the crash occurred. In this study, the majority of crash records were collected from MnCMAT database. The crash time information mostly originated from the police reports. However, Hourdos (2007) found that the crash time reported by police was “second-hand” information, which was not accurate to be considered as “real” crash time due to many reasons. There is no value to analyze the traffic conditions prior to the crashes if the use of crash time is not authentic. Hence, crash times used in this study were estimated rather than using the reported times.

The methodology used in estimating crash time was the basic application of shockwave in traffic flow theory (Elefteriadou, 2014). Crash times were estimated based on the time when obvious short-time fluctuations in traffic were detected at upstream or downstream detectors, and the travel time of shockwave potentially caused by the fluctuation of traffic.

$$T_c = T_L - t_w \quad (3.1)$$

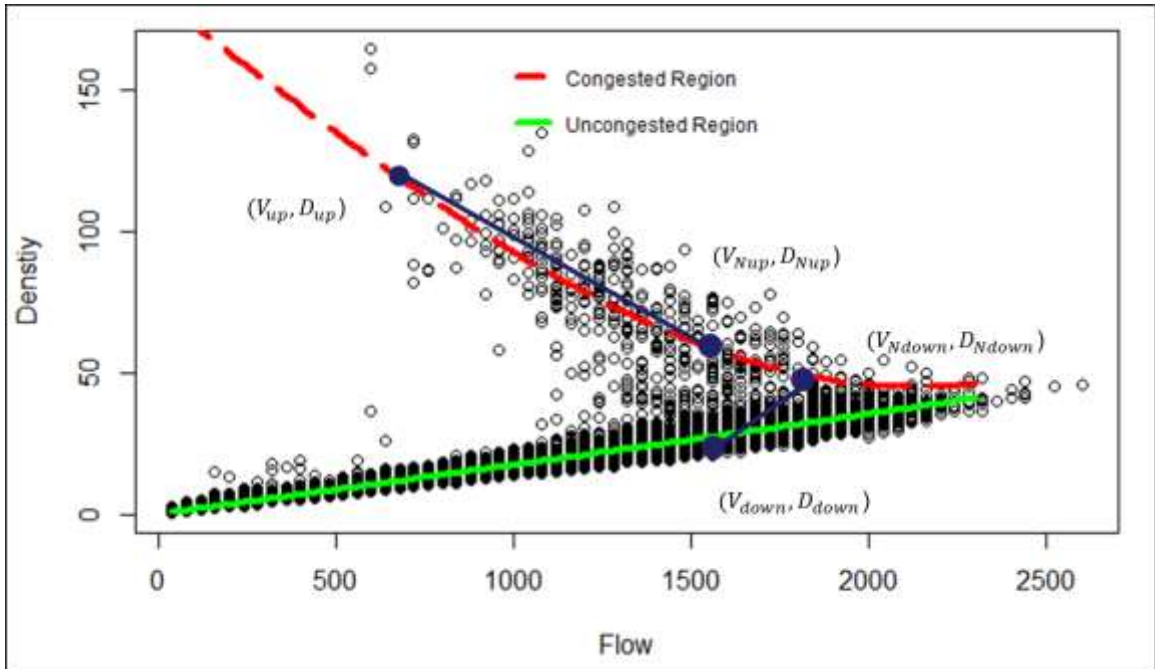
Where

$T_c$ = estimated crash time;

$T_L$ = the time when obvious traffic fluctuation is detected at loop detectors; and

$t_w$ = shockwave travel time.

To obtain the shockwave travel time, it is sufficient to derive the shockwave propagation speed, because the distance between the crash and loop detectors is a known factor based on their locations. For each VMS impact region where crash occurred, there were two pairs of loop detectors selected at upstream and downstream for traffic detection.



**FIGURE 5 Flow-density Model for VMS Site V94E08**

First, a flow-density model was constructed for estimating shockwave propagation speed. One of the simple models in traffic flow theory is the Greenshield's model (Roess et al. 2004), which builds up the relationships among traffic flow, speed, and density under the assumption of a linear speed-density relation. In this study, each VMS impact area had its unique flow-density model, and the construction of model was separated into

two subset models: one for congestion region, the other for free flow traffic region. Linear regressions were used for both models with ordinary least squares method. (FIGURE 5 shows one of the examples) Flow and density data were extracted on 10 normal weekdays with 30-second time intervals at the corresponding detectors. For each model, density was the function of flow. This is because that measured volume was relatively more reliable than measured occupancy, which was used for calculating density. Occupancy is the proportion of time that a detector is "occupied," or covered, by a vehicle in a defined period of time (Roess et al. 2004). For all the detectors in the study regions, magnetic detectors were used to measure occupancy and volume, and the measurement of occupancy sometimes biased on size of vehicles. Generally, a small vehicle causes greater increase in occupancy than larger vehicles, because it takes less time for magnetic field of loop to detect the metal surfaces on the under carriage. Measured occupancy is not accurate in the region where frequent lane changing and giant trucks passing are prevalent.

With measured flow and estimated density, shockwave propagation speed could be estimated.

$$\text{Upstream: } S_{w\_Up} = \frac{V_{up} - V_{N_{up}}}{D_{up} - D_{N_{up}}} \quad (3.2)$$

$$\text{Downstream: } S_{w\_Down} = \frac{V_{down} - V_{N_{down}}}{D_{down} - D_{N_{down}}}$$

Where

$S_{w\_Up}$  = estimated shockwave propagation speed to upstream;

$S_{w\_Down}$  = estimated shockwave propagation speed to downstream;

$V_{up}$  ( $V_{down}$ ) = median measured flow (5 minutes with 30-second interval) after traffic breakdown at upstream (downstream) detectors;

$D_{up}$  ( $D_{down}$ ) = estimated density after traffic breakdown at upstream (downstream) detectors;

$V_{N_{up}}$  ( $V_{N_{down}}$ ) = median measured flow (10 minutes with 30-second interval) before traffic breakdown at upstream (downstream) detectors; and

$D_{N_{up}}$  ( $D_{N_{down}}$ ) = estimated density before traffic breakdown at upstream (downstream) detectors.

Crash times were estimated using shockwave propagation speed, which was calculated with upstream traffic conditions or downstream conditions. Ideally, the gap between the results based on both conditions should be slight. However, large gaps were detected sometimes. Generally, crash time estimated by upstream detectors was more reliable than the one estimated by downstream detectors. The change of occupancy at upstream detectors was more easily detected when most of crashes occurred. In addition, traffic conditions at downstream were less predictable and vary by the impact of the crashes and levels of congestion. Thus, estimated crash time was defined as the time estimated based on upstream detectors, if a gap was greater than two minutes. However, if the gap was less than two minutes and the crashes were close to the downstream detectors, downstream traffic conditions contributed to the crash time estimation. In this case, a weighted estimated crash time was used. The calculation of estimated crash time is showed in Equation (3.3). FIGURE 6 illustrates the conditions for the estimation.

$$T_C = \begin{cases} T_{C_{up}}, & \text{if gap} > 2 \text{ mins} \\ T_{C_{up}} \left( \frac{L_{down}}{L_{up} + L_{down}} \right) + T_{C_{down}} \left( \frac{L_{up}}{L_{up} + L_{down}} \right), & \text{if gap} < 2 \text{ mins} \end{cases} \quad (3.3)$$

And

$$T_{C_{up}} = (T_{L_{up}} - \frac{L_{up}}{|S_{w_{up}}|}) \quad (3.4)$$

$$T_{C_{down}} = (T_{L_{down}} - \frac{L_{down}}{|S_{w_{down}}|}) \quad (3.5)$$

Where

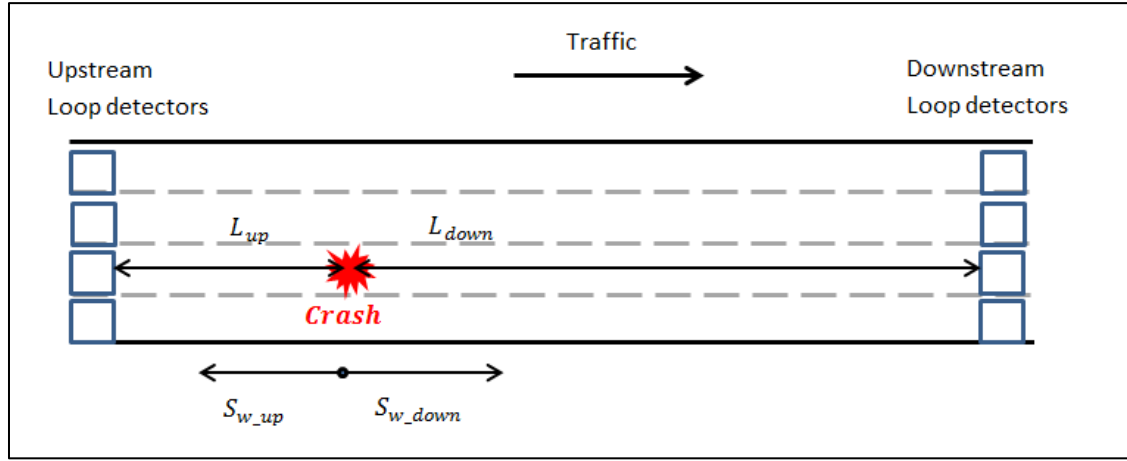
$T_C$  = estimated crash time;

$T_{C_{up}}$  ( $T_{C_{down}}$ ) = estimated crash time based on upstream (downstream) detectors;

$L_{up}$  ( $L_{down}$ ) = the distance between the crash and upstream (downstream) detectors;

$T_{L_{up}}$  ( $T_{L_{down}}$ ) = the time when obvious traffic breakdown was detected at upstream (downstream) detectors; and

$|S_{w\_up}|$  ( $|S_{w\_down}|$ ) = absolute value of shockwave propagation speed to upstream (downstream) detectors.



**FIGURE 6 Illustration of Crash Time Estimation**

Validation was conducted after crash times were estimated. The results would become suspicious if there was a huge gap between the estimated crash time and reported crash time (greater than 20 minutes). One of the reasons for huge gaps may result from the imprecise use of crash locations. The information of crash locations in the estimation process was extracted from the police report (MnCMAT), which was difficult to validate. Another reason was that some “unknown events” were blended in the traffic data when crashes occurred, and they were not easy to detect based on the existing data. However, a crash record was removed from the database if the crash time could not be estimated based on the existing information.

In summary, crash time estimation is crucially necessary before collecting traffic data. The estimation methodology stated above minimized the bias of the improper use of crash times, and it was applied to all the available crashes.

### **3.4.2 Traffic Conditions Prior to the Crashes**

The crash-prone traffic conditions are defined as traffic conditions two minutes prior to the estimated crash time at upstream and downstream detectors. The basis for determining the time interval between the starting time of data collection and the

estimated crash time is the shockwave travel time. The time interval should not be longer than the shockwave traveled between downstream and upstream detectors. The reason for this argument is because, only the shockwave that could potentially lead to the crashes that need to be captured. Shorter or longer time intervals can both result in bias of the results by either losing traffic information, or capturing a shockwave that is irrelevant to crash. The empirical shockwave propagation speed of two congested states was around 14 to 20 mph (Chung, 2011; Wu and Liu, 2011), which was consistent with the results estimated in this study. The distance of adjacent loop detectors in Minnesota city freeway system is roughly between 0.4 mile and 0.6 mile. Hence, the time for one shockwave to travel from downstream to upstream is 106 seconds for the average of shockwave propagation speed and spacing of detectors. Thus, the time interval for extracting traffic condition data prior to the crashes was up to two minutes, in order to avoid capturing more than one shockwave.

**TABLE 6 Traffic Metric list and Description**

Traffic metrics	Description
AvgS_up	Upstream average speed
MAXDiff_S_up	Upstream maximum speed difference of adjacent lanes
COV_S_up	Upstream coefficient of variance in speed
AvgOcc_up	Upstream average occupancy
Flow_up	Upstream traffic flow
StdS_up	Upstream standard deviation of speed
StdO_up	Upstream standard deviation of occupancy
AvgS_down	Downstream average speed
MAXDiff_S_down	Downstream maximum speed difference of adjacent lanes
COV_S_down	Downstream coefficient of variance in speed
AvgOcc_down	Downstream average occupancy
Flow_down	Downstream traffic flow
StdS_down	Downstream standard deviation of speed
StdO_down	Downstream standard deviation of occupancy



Traffic conditions prior to crashes can be expressed in various traffic metrics. According to previous studies stated in Chapter 2, the upstream and downstream variation of speed showed an association with crashes. Moreover, levels of congestion also had impact on crash occurrence. The traffic metrics used in this study are shown in TABLE 6. All of the traffic metrics were derived based on the two-minute time interval rule. The process was applied to all the 183 crash records extracted from MnCMAT.

The basic measurements utilized for deriving traffic metrics stated in TABLE 6 are traffic flow, occupancy, and speed. In this study, the collection of those measurements relied on MnDOT's data extraction tool. This tool is developed based on Java platform, and it enables users to download basic traffic flow characters for each loop detector at any time of year for various time intervals.

For each loop detector, only volume and occupancy data is measured. (Some detectors also measure speed, but this was not used in this study.) Flow data is calculated by volume; headway and density are derived from occupancy together with detector width; and speed is estimated based on volume and occupancy. Although speed data used in this study was not precisely measured, the majority of values estimated by this tool were reasonable and adequate to reflect the trend of speed for each site. Unreasonable speed data, for example speeds over 100 mph, were detected and removed from the database with an excel macro developed in this study.

In summary, traffic flow, occupancy, and speed data two minutes prior to the estimated crash time can be downloaded directly using MnDOT data tools. All the traffic metrics were derived based on those basic measurements. Ideally, for each crash record, data for traffic conditions were available after all the traffic metrics stated in TABLE 6 were derived. However, due to the availability of the detector data and the location of the loops, upstream traffic data for VMS site "V94W07" were not reasonable in most cases, for example, average speed was generally over 90 mph. In addition, traffic data for VMS site "V94W07" was not accessible.

### **3.5 Weather Data Collection**

#### **3.5.1 Weather data Sources**

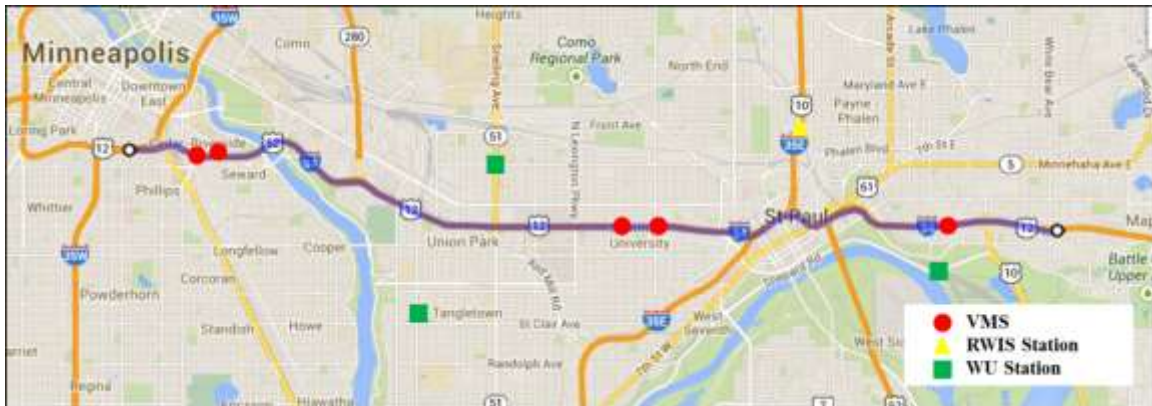
Weather data in the study were available from two data sources: one was Road Weather Information System (RWIS), and the other was Weather Underground. The main types of weather information used in this study were precipitation types (rain, snow, etc.) and precipitation rates. Time range for weather data collection was from January 1, 2006 to December 31, 2012, which was corresponding to the crash and VMS data.

Road Weather Information System (RWIS) is a central system mainly operated by Federal Highway Administration (FHWA) to collect field data from numerous environmental sensor stations (ESS), and it serves as a communication system for data transfer. Three types of road weather data could be accessed from the RWIS system: atmospheric data, pavement data, and water level data. Precipitation type, precipitation rate, and temperature in the first type of the database were critical attributes. RWIS historical weather data in Minnesota were downloaded from SCAN Web operated by MnDOT. Weather Underground is a commercial weather service that offers free real-time and historical weather information to the public. Weather historical data from various weather stations in Minnesota metro area can be accessed from the “Historical Weather” option at Weather Underground official website.

#### **3.5.2 Advantages and limitations**

The main advantages of RWIS are that it provides historical weather information with the time interval around 5 minutes, and various weather-related attributes can be collected. Moreover, the average precipitation rate is computed every minute, which could guarantee the accuracy of the data. However, one of the disadvantages is, RWIS has limited weather stations in the Minnesota Metro area, especially in the study regions. In addition, the problem of data missing for some weather stations is another issue. Compared to RWIS, Weather Underground has several weather stations close to the study regions. However, limited type of information can be downloaded from Weather Underground. The precipitation type, for example, is not available if the selected time interval for weather data is less than one hour. In addition, the average time intervals of

historical weather data in Weather Underground are around 8 minutes, which is longer than RWIS data.



**FIGURE 7 Potential Weather Stations Selected (Screenshot from Google Maps)**

### 3.5.3 Weather Stations Selection and Data Validation

Based on the locations of VMS impact regions, four stations (one from RWIS, and three from Weather Underground) were selected as potential sites for downloading weather data. The locations of stations are displayed in FIGURE 7, and their distances to VMS sites and availability of data are summarized in TABLE 7. According to the data availability table, RWIS has much more data missing than Weather Underground, and those missing data cannot be replaced, because other RWIS stations are more than 10 miles away from the study regions. On the contrary, Weather Underground has several stations available around the study areas. In order to validate weather data from the two data sources, historical annual precipitation data was downloaded from the website of Minnesota Climatology Working Group (MCWG), which is managed by Minnesota State Climatology Office, during year 2006 to year 2012. In comparison, cumulative precipitation rates reported by Weather Underground were more consistent with MCWG than RWIS data, and the data were more reasonable in total.

### 3.5.4 Weather Data Used in this Study

In this study, three major weather stations from Weather Underground were selected for data collection. The stations information corresponding to the VMS impact regions is

summarized in TABLE 8. Substitute stations were stations that could provide alternative weather data in case data in major stations were missing. In total, a seven-year period of weather data was downloaded with the time interval less than 15 minutes. The raw data included time, temperature, dew point, humidity, wind speed, gust, pressure, and precipitation. The critical information needed was precipitation rates and type.

**TABLE 7 Potential Weather Station Locations and Data Availability**

Source	Station	Miles to W05	Miles to W07	Miles to W09	Miles to E08	Miles to E11	Data missing (2010 to 2012, %)
RWIS	Cayuga St. Bridge	2.16	2.04	7.34	7.15	2.49	27.8
WU	Blair Manor	5.69	2.17	3.58	3.37	1.7	4.37
WU	Groveland Park	6.7	3.18	3.32	3.17	2.72	14.5
WU	Mounds Park	0.51	3.52	9.16	8.98	4.02	12.9

**Notes:** WU is short for Weather Underground; “W05” is short for “V94W05”, “W07” is short for “V94W07”, and so on.

**TABLE 8 Weather Stations Used in this Study**

VMS Site ID	Corresponding Stations and Downloading Time Period	Substitute Stations
V94W05	Mounds Park (2006-2012)	Groveland Park, Blair Manor, St. Paul, Saint Paul MN US
V94W07	Groveland Park(2006-2008), Blair Manor (2009-2012)	Mounds Park, U of MN
V94W09	Groveland Park(2006-2008), Blair Manor (2009-2012)	Mounds Park, U of MN
V94E08	Groveland Park(2006-2008), Blair Manor (2009-2012)	Mounds Park, U of MN
V94E011	Groveland Park(2006-2008), Blair Manor (2009-2012)	Mounds Park, U of MN

However, precipitation type information was not available in the historical reports with time interval less than 1 hour. The way to address the issue was that, if precipitation rates with 5-minute time interval, for example, were detected to be greater than zero, the

precipitation types were inferred from the “conditions” information based on 1-hour time interval. The “conditions” generally included cloudy, rain, snow, etc. RWIS data was also used for validation if Weather Underground data was missing or ambiguous.

## **Chapter 4: Impact of VMS on Short-Time Change of Traffic Speed**

According to the previous studies, speed decreases were detected when a VMS was activated, especially when there was a safety concern stated in the messages, or the message information was overloaded. (Haghani, 2013) In addition, many researchers believed that variation of speed was a causal factor for crashes. (Oh et al., 2001; Hourdos et al., 2007; Abdel-Aty and Pande, 2004) Hence, the study of VMS impact on short-time traffic fluctuations is critical for the safety evaluation. Speed change before and after the activation or deactivation of VMS has been widely studied in recent years. However, seldom considered were other causal factors that could cause changes in speed in VMS impact regions, rather than the display of sign messages. Moreover, in previous studies, most of traffic data used was highly aggregated, and there were few control variable presented in the analysis.

In this study, a generalized linear model (Dobson and Barnett, 2008) was constructed to investigate the change of speed when VMS was activated and deactivated, using disaggregated time of day traffic data. In addition, weather factors were used in the analysis in order to test whether the impact of VMS was associated with adverse weather conditions.

### **4.1 Data Preparation**

#### **4.1.1 VMS Logs Data**

Four VMS logs in Year 2011 were used in this study (See TABLE 9). In order to capture the change of speed, the selection of VMS devices requires the distance between an upstream detector and the VMS panel should be within the vision field of drivers, which is greater than or equal to 860 feet (Harder et al., 2003).

The time when a VMS was activated and deactivated was the critical information for this study. The attribute “description” in the logs describes the status of a VMS as “Sign DEPLOYED” or “Sign CLEARED”. The event time associated with “Sign DEPLOYED” begins “active time”; and the event time associated with “Sign CLEARED” begins “inactive time”. One of three time period statuses, “AM Peak”, “PM Peak” and “Off Peak”, were assigned to each “active time” or “inactive time” for evaluating the impact of

VMS during different times of day. According to the Twin Cities Metropolitan Area Freeway and Major Arterials Crash Summary, prepared by the MnDOT Regional Transportation Management Center (RTMC, 2011), the AM Peak was defined as 6:00 AM to 8:59 AM, and PM Peak was 2:00 PM to 6:59 PM, Monday through Friday. Off Peak was for non-peak weekday hours and weekends. In total, 578 “active” cases and 542 “inactive” cases were extracted from four VMS logs in Year 2011.

**TABLE 9 VMS Sites Selection for the Study of VMS Impact on Speed**

VMS ID	Cross Street	Detectors ID	Distance between detectors and VMS (ft)
V94E08	I-94@25th Ave	S554	840
V94E11	I-94@ Victoria St	S778	90
V94W05	I-94@ Earl St	S1070	158
V94W09	I-94@25th Ave	S553	528

#### 4.1.2 Traffic Speed Data

MnDOT’s data extraction tool was utilized to download traffic data, and an Excel Macro was applied to organize the data. Four groups of detectors were selected based on the locations of available VMS devices (See TABLE 9). In the study regions, there are three or four lanes available for each segment, but the speeds might not be consistent for all lanes. Typically, the rightmost lane has lower occupancy and higher variation of speed. However, the impact of VMS is supposed to be same for all lanes at each segment. Hence, a weighted average speed (Equation 4.1) was used. Traffic volume was used for calculating the weights. Speed and volume data were downloaded in 30 second intervals for 24 hours a day, and 365 days in 2011 for each group of detectors. However, the limitation of using loop detector data is that speeds were estimated rather than being measured, and do not reflect individual vehicle actions.

Weighted average of speed is

$$\bar{s} = \frac{\sum_{i=1}^k (s_i \times V_i)}{\sum_{i=1}^k V_i} \quad (4.1)$$

Where

$\bar{s}$  = weighted average of speed;

$s_i$  = speed in lane  $i$ ;

$V_i$  = volume in lane  $i$ ; and

k = number of lanes at the road section.

#### 4.1.3 Weather Data

Weather data were collected from stations of Weather Underground (See TABLE 8). Weather conditions were linked to the each case by matching the time of event.

#### 4.1.4 Data Combination and Filtering

In total, 578 “active” cases and 542 “inactive” cases from 4 VMS impact regions were available. For each case, speeds for 30-second time intervals in consecutive 5 minutes before and after the message “active time” or “inactive time” were extracted. Then the average speed before ( $x_{bi}$ ) and after ( $x_{ai}$ ) the change of VMS status was calculated; and the change of speed ( $\delta_i$ ) was derived (Equation 4.2). Based on the VMS activation or deactivation time, weather conditions were linked to each case. In sum, for a single event, weather conditions and average speed change in the VMS impact region were available corresponding to the status of VMS (active and inactive), and the time of the day (AM Peak, PM Peak, and Off Peak) when a message was displayed.

$$\begin{aligned} x_{bi} &= \frac{\sum_{j=1}^{10} \overline{s_{bij}}}{10} \\ x_{ai} &= \frac{\sum_{j=1}^{10} \overline{s_{aij}}}{10} \\ \delta_i &= \overline{x_{ai}} - \overline{x_{bi}} \end{aligned} \tag{4.2}$$

Where

$x_{bi}$  = average speed before the change of VMS status for case  $i$ ;

$x_{ai}$  = average speed after the change of VMS status for case  $i$ ;

$\overline{s_{bij}}$  = weighted average speed in  $j^{\text{th}}$  30-second before the change of VMS status for case  $i$ ;

$\overline{s_{aij}}$  = weighted average speed in  $j^{\text{th}}$  30-second after the change of VMS status for case  $i$ ;

and

$\delta_i$  = average change of speed for case  $i$ .

However, the VMS is not the only factor that can influence speed. Downstream shockwaves caused by the events described by the sign messages, for example, is one of



the potential variables associated with the speed change in VMS impact regions. To erase the noise of shockwaves and guarantee the authenticity of the results to the maximum extent, those cases with potential downstream shockwaves that could reach the VMS impact regions within the analysis time range need to be removed from the analysis.

Downstream shockwaves were identified if a VMS was activated because of the events, which were described in the sign messages, could potentially produce shockwaves to the upstream detectors. In order to determine whether a shockwave could reach the VMS impact regions, and affect the speed, two factors were considered: distance from the source of shockwaves to the VMS impact regions ( $L$ ), and the shockwave propagation speed ( $s_w$ ). The location of the potential shockwave sources could be roughly developed from VMS messages. Typically, the second line of messages is the events location statement, for example “AT HYW 280”. Thus, the distance ( $L$ ) was measured based on the locations of VMS and potential shockwave sources. The shockwave propagation speed ( $s_w$ ) between two traffic states was assumed 15 mph in this section.

The critical shockwave travel time ( $t_0$ ) is defined as minimum allowed shockwave travel time. Since speed was extracted for 5 minutes before and after a VMS was activated, and the reaction time for VMS operators was 2 minutes on average,  $t_0$  was set as 7 minutes in this study. If shockwaves were not able to arrive at VMS impact regions within the critical travel time ( $t_0$ ), equation (4.3) below would be satisfied.

$$\frac{L}{s_w} > t_0 \xrightarrow{\text{Transform}} L > s_w \times t_0 \quad (4.3)$$

The critical distance ( $L_0$ ) is defined as distance that the potential downstream shockwaves traveled during the critical travel time  $t_0$  (FIGURE 8). It is the minimum distance between the source of shockwaves and the site of VMS for the VMS impact regions to be unaffected by downstream shockwaves (Equation 4.4).

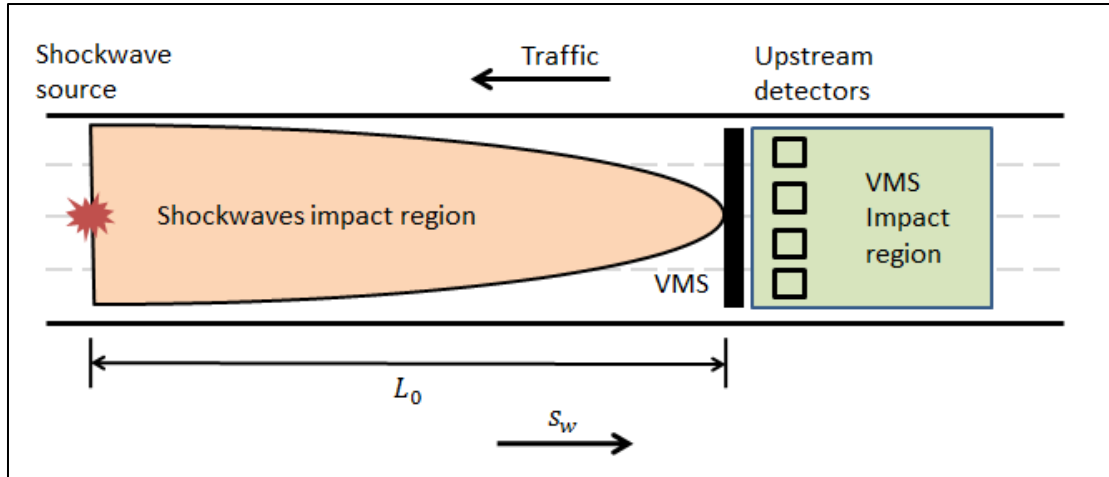
$$L_0 \approx s_w \times t_0 = 15 \times \frac{7}{60} = 1.75 \text{ miles} \quad (4.4)$$

Where,

$L_0$  = critical distance;

$s_w$  = shockwave propagation speed (15 mph);

$t_0$  = critical shockwave travel time (7 minutes).



**FIGURE 8 Shockwave Critical Distances to VMS Impact Regions**

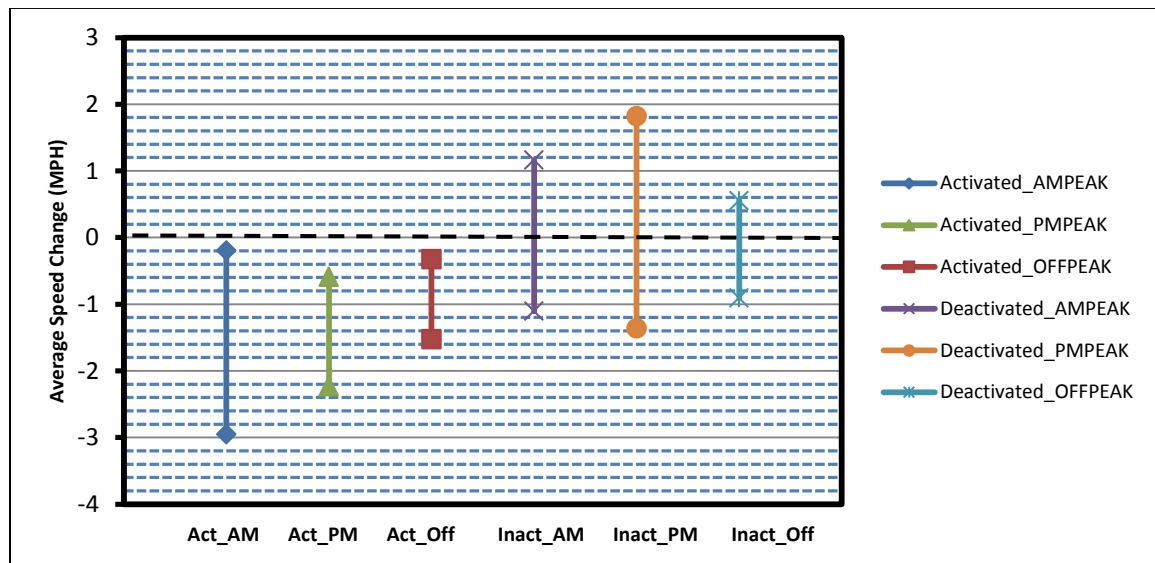
In sum, a case was removed from the study if the corresponding VMS message indicated the event downstream could potentially produce shockwaves, and the distance from the source of shockwave to VMS impact regions ( $L$ ) was less than the critical distance ( $L_0$ ). The potential impact events were generally non-recurring incidents such as crash, stalled vehicles, and snow removal. Moreover, if sign messages indicated a potential downstream event but had ambiguous location, such as “CRASH LEFT LANE”, the corresponding case was also removed from analysis.

However, it is more difficult to identify shockwaves associated with VMS deactivation. According to the engineers in RTMC, signs are deactivated only after the operators confirm the events are clear from the cameras or get verification from the police officers. The “inactive time” cannot reflect the impact of the events as accurately as the “active time”. Hence, “inactive” cases were removed if the corresponding messages indicated the previous events were likely to produce shockwaves that could potentially affect the speed in the VMS impact regions. Those events primarily included crashes and incidents that occurred in the freeway main lanes. After removing events potentially confounded by

shockwaves, 670 cases (403 “active” cases and 267 “inactive” cases) were available for further analysis.

## 4.2 Model Construction and Methodology

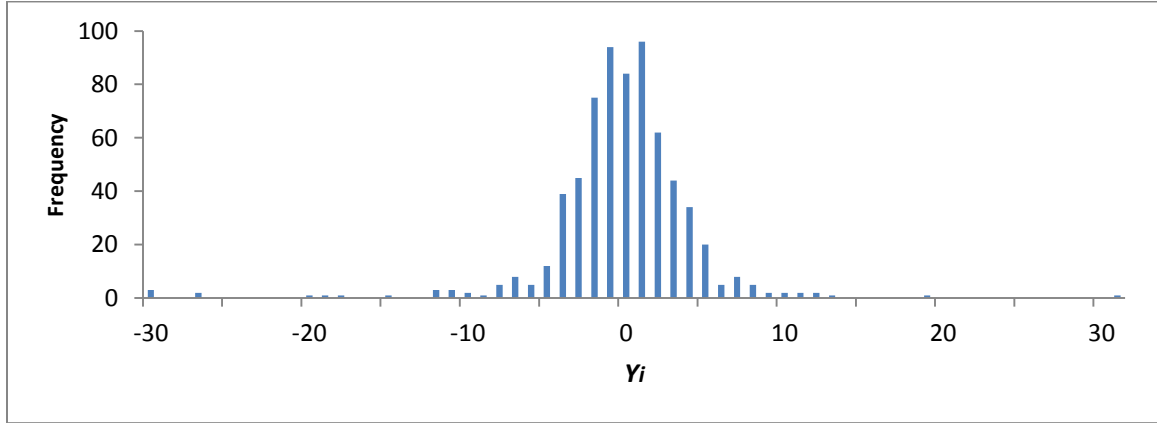
Normal linear models were used to investigate the VMS impact on short-time speed change. To be better understanding the mechanism of the effect, cases that had potentially been influenced by downstream shockwaves were removed from the analysis. The association between the deployment of the VMS and speed change was studied under different times of the day (AM Peak, PM Peak, and Off Peak), VMS statuses (active, and inactive), and weather conditions (normal weather, rain, and snow).



**FIGURE 9 95% Confidence Intervals for the Average Speed Change**

Under the assumption that the impact of VMS on change of speed is similar for all the four impact regions selected in this study, the 95% confidence intervals of average speed change for each time range of day and different VMS statuses were calculated and plotted in FIGURE 9. Speed decreases were observed when VMS was activated for each time of the day. The drop of speed was slightly higher in peak hours than during off-peak hours. However, a change of average speed when VMS was deactivated was not apparent. Especially in off-peak hours, the speed change was within 1.0 mph, which was negligible

for practical purpose. The plot of 95% confidence intervals of average speed change provided the trend of variation of speed when a VMS was activated or deactivated, but the other factors were unknown. A normal linear model was constructed to address the issue.



**FIGURE 10** Frequency of Average Speed Change

Supposing  $Y_i$  represents the change of speed for case  $i$ , which is the difference of speed before ( $x_{ai}$ ) and after ( $x_{bi}$ ) a VMS change in status, and assuming that  $Y_1, Y_2, \dots, Y_N$  are independent random variables. The distribution of  $Y$  is approximately close to normal distribution based on the histogram plots of  $Y$  (See FIGURE 10). In this case, a generalized normal linear model (Dobson and Barnett, 2008) was constructed with the form,

$$E(Y_i) = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}; \quad Y_i \sim N(\mu_i, \sigma^2). \quad (4.5)$$

The expectation of  $Y_i$  is

$$\mu_i = \beta_0 + \sum_{j=1}^{3-1} \beta_{1j} T_{1j} + \sum_{k=1}^{2-1} \beta_{2k} S_{2k} + \sum_{m=1}^{3-1} \beta_{3m} W_{3m} \quad (4.6)$$

Where

$Y_i$  = average change of speed;

$T_{1j}$ = time of day factor,  $j=1, 2, 3$ ;  
 $S_{2k}$ = VMS status factor  $k=1, 2$ ;  
 $W_{3m}$  = weather factor.  $m=1, 2, 3$ ; and  
 $\beta$ = coefficients.

$$T_i = \begin{cases} 1, & \text{Case occurs in the Time of Day}_p \\ 0, & \text{Others} \end{cases} \quad (4.7)$$

$$S_i = \begin{cases} 1, & \text{Case occurs in the VMS Status}_p \\ 0, & \text{Others} \end{cases}$$

$$W_i = \begin{cases} 1, & \text{Case occurs in the Weather conditions}_p \\ 0, & \text{Others} \end{cases}$$

\*Notes: Time of Day includes AM Peak, PM Peak, and Off Peak; VMS Status includes VMS activated, and VMS deactivated; weather conditions include rain and snow.

The least-squares estimator of  $\beta$  is

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (4.8)$$

under the condition that  $X^T X$  is non-singular. The factors for time of day, VMS status, and weather conditions are discrete, nominal scale variables (Equation 4.7), and the numbers of the variable have no numeric significance. A collection of dummy variables was generated by R when fitting the models to avoid singular explanatory variables. In order to test the change of speed in different situations, a base condition was defined. For the convenience of interpretation, the base conditions should have the characteristics that the impact of VMS on average speed change was minimized. By looking at FIGURE 9, the change of average speed in off-peak hours, when VMS was deactivated, was the least recognizable among all the combinations. Thus, we define it as the base condition (Condition 1) for scenario one, for which the variables for “Off Peak” and “VMS

Deactivation” were removed from model (4.6). Similarly, the second scenario was assumed to have the base condition that VMS was activated in the off-peak hours (Condition 7). Both base conditions were assumed to occur in normal weather conditions. The two scenarios and the test conditions are summarized in TABLE 10.

**TABLE 10 Scenarios for Testing the Change of Average Speed under Different Conditions**

Scenario one				Scenario two			
Condition ID	Time of day	VMS status	Weather	Condition ID	Time of day	VMS status	Weather
1	Off Peak	Deactivated	None	7	Off Peak	Activated	None
2	AM Peak	Deactivated	None	8	AM Peak	Activated	None
3	PM Peak	Deactivated	None	9	PM Peak	Activated	None
4	Off Peak	Activated	None	10	Off Peak	Deactivated	None
5	Off Peak	Deactivated	Rain	11	Off Peak	Activated	Rain
6	Off Peak	Deactivated	Snow	12	Off Peak	Activated	Snow

### 4.3 Analysis of VMS Impact

The software package, R, was used to fit models and test hypotheses. R is a language and environment for statistical computing and graphics, and it has been widely used in statistical practice. As stated above, the analysis was conducted for two scenarios.

$$E(Y_i) = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}; \quad Y_i \sim N(\mu_i, \sigma^2) \quad (4.9)$$

$$\text{Scenario one:} \quad \mu_i = \beta_0 + \beta_1 t_{ai} + \beta_2 t_{pi} + \beta_3 S_{ai} + \sum \beta_4 W_i \quad (4.10)$$

$$\text{Scenario two:} \quad \mu_i = \beta_0 + \beta_1 t_{ai} + \beta_2 t_{pi} + \beta_3 S_{ina i} + \sum \beta_4 W_i \quad (4.11)$$

Where

$Y$  = average change of speed;

$t_{ai}$  = AM Peak factor;

$t_{pi}$  = PM Peak factor;

$S_{ai}$  =active VMS factor;

$S_{inai}$  =inactive VMS factor;

$W_i$  = weather factor; and

$\beta_i$ = coefficients

#### 4.3.1 The Impact of VMS in Adverse Weather Conditions

Weather conditions are generally supposed to affect freeway driving. However, after running both scenarios in R, there was no evidence that the coefficients for rain and snow factors were significantly different from zero. In addition, when both weather variables were removed from the models, the change of coefficients and t-values was not noteworthy for the other variables. (See TABLE 11) However, there is an argument that the sample sizes for rain variable (24 cases from 670) and snow variable (4 cases from 670) were not sufficient. For instance, there was only one sample for the condition that a VMS was activated in off-peak hours under snow weather. Thus, a combination of adverse weather variable was more desirable than analyzing each variable separately in the model.

**TABLE 11 Regression Results With and Without Weather Variables Comparison (Scenario One Example)**

	Variable	Coefficient	Std.Error	t-value	p-value	Significance
	(Intercept)	-0.026	0.320	-0.082	0.935	
Model with weather variables	$t_{ai}$	-0.311	0.517	-0.602	0.547	
	$t_{pi}$	-0.171	0.479	-0.356	0.722	
	$S_{ai}$	-1.056	0.386	-2.740	0.006	**
	$W_{Ri}$	0.760	1.005	0.756	0.450	
	$W_{Si}$	-2.260	2.418	-0.935	0.350	
Model without weather variables	(Intercept)	-0.018	0.315	-0.057	0.954	
	$t_{ai}$	-0.296	0.517	-0.572	0.568	
	$t_{pi}$	-0.209	0.477	-0.438	0.662	
	$S_{ai}$	-1.038	0.385	-2.699	0.007	**

**Notes: Significance codes: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, 0.1 ‘ ’ 1**

In order to investigate whether it was appropriate to combine the adverse weather variables, an analysis of variance (ANOVA) was performed, and results are displayed in TABLE 12. The null hypothesis was that the model “A” was not significantly better than the model “B<sub>i</sub>”. First, if the model without weather variables (model A) was compared to

the model with separated weather variables ( $B_1$ ), the p-value, which was  $4.28e^{-8}$  (smaller than the significance level  $\alpha=0.05$ ), indicated that the null hypothesis was rejected. However, a p-value of 0.09 from the ANOVA result when comparing model A and model  $B_2$  revealed that the null hypothesis was accepted. Since there was no significant difference between model A and model  $B_2$ , the weather variables were appropriate to be combined. Thus, instead of including “snow” and “rain” variables in the models, the combined weather variable was named as “bad weather” ( $W_{Bi}$ ), and was utilized in the analysis of both scenarios.

**TABLE 12 Model Selections (Scenario One Example)**

Models		Analysis of Variance (ANOVA)		
ID	Description	Difference DF	Difference Dev	p-value
A	No Weather Variable			
$B_1$	With Separated Weather Variables	2	33.932	$4.28e^{-8}$
$B_2$	With Combined Weather Variables	1	2.832	0.09

**Notes:** DF is degree freedom; Dev is deviance.

#### 4.3.2 Analysis of Difference Scores

Generalized linear models were used to analyze the change of speed affected by the displaying of sign messages. The explanatory variables included disaggregated time of day factors, VMS status factors, and weather factors. However, the initial speed was not considered in both scenarios, since the response was a function of it. Its influence on the outcomes was analyzed in the next section. In terms of the analysis of difference scores, speed changes were estimated in 10 conditions (rather than 12 conditions since weather variables are combined), and the significance of change was tested.

##### *a. Scenario One*

The base condition for scenario one was that VMS was deactivated in off-peak hours under normal weather conditions (Equation 4.10). The regression results given by R were displayed in the TABLE 16 (Different Scores Model). R was able to provide the values of coefficients ( $\beta_i$ ), the standard error ( $s_i$ ), t-statistics ( $t_i$ ), the corresponding p-values ( $p_i$ ), and covariance matrix ( $\sigma_{ij}$ ).

For the base condition (Condition 1, a VMS was deactivated in off-peak hours when weather condition was normal), the estimated change of average speed is  $\beta_0$ , which is the



intercept of the model results. For condition  $k$  ( $k=1, 2, 3, \dots$ ), the change of average speed is

$$\beta_0 + \beta_k ,$$

where  $\beta_k$  is the coefficient of the corresponding variable in condition  $k$ . The null hypothesis is,

$$H_0: \beta_0 + \beta_k = 0 , \quad (4.12)$$

The alternative hypothesis is,

$$H_A: \beta_0 + \beta_k \neq 0 ,$$

Z-test was used for testing the hypothesis. When  $n$  is great enough (670 total observations in this study), the test statistic is asymptotic standard normal distributed, and the test statistic is,

$$Z = \frac{\beta_0 + \beta_k - 0}{\sqrt{\sigma_1^2 + \sigma_k^2 + 2\sigma_{1k}}} \quad (4.13)$$

The test for significance was performed under five conditions at significance level of  $\alpha = 0.05$ . Results presented in TABLE 13 revealed that the average change of speed was not significant when VMS was deactivated under all conditions. Time of day factors and weather factor did not affect the speed change. These outcomes indicated that drivers did not tend to accelerate or decelerate when a sign was turned off on any time of the day and any weather conditions. However, compared to the change when a VMS was deactivated, a decrease was detected when a VMS was activated in off-peak hours (Condition 4), but the magnitude of decrease was only about 1.0 mph. The model results were consistent with the 95% confidence interval plots in FIGURE 9.

**TABLE 13 Significance Test for the Average Speed Change in Scenario One**

Condition ID	Time of day	VMS Status	Weather	Speed Change	Z-statistic	p-value	Significance
1	Off Peak	Deactivated	None	-0.04	-0.116	0.907	
2	AM Peak	Deactivated	None	-0.33	-0.638	0.523	
3	PM Peak	Deactivated	None	-0.23	-0.463	0.643	
4	Off Peak	Activated	None	-1.07	-3.637	$2.76e^{-4}$	***
5	Off Peak	Deactivated	Bad	0.29	0.310	0.757	

**Notes: Significance codes: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, 0.1 ‘ ’ 1**

*b. Scenario Two*

For scenario two, the base condition was converted to the situation that VMS was activated in off-peak hours under normal weather conditions. The model (4.11) was the same as in scenario one except the variable for VMS status ( $S_i$ ) was switched to the “inactive” ( $S_{inai}$ ). The regression results (TABLE 14) were generated by R.

**TABLE 14 Regression Result for Scenario Two**

Explanatory Variable	Coefficient	Std. Error	t-value	p-value	Significance
(Intercept)	-1.072	0.295	-3.637	$2.97e^{-4}$	***
$t_{ai}$	-0.294	0.517	-0.568	0.570	
$t_{pi}$	-0.195	0.479	-0.407	0.684	
$S_{inai}$	1.034	0.385	2.685	0.007	**
$W_{Bi}$	0.326	0.933	0.350	0.727	

**Notes: Significance codes: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, 0.1 ‘ ’ 1**

In scenario two,  $\beta_0$  (intercept) described the average change of speed under the condition that VMS was activated in off-peak hours with normal weather conditions (TABLE 14). The change of speed (-1.07 mph) was significant based on the p-value, and it was consistent with the result from scenario one. The procedures for testing the significance of change under other conditions were conducted the same way as in scenario one.

Analysis outcomes (TABLE 15) revealed that speed decreased when the weather condition was normal at any times of the day. However the reductions of speed were less than 2.0 mph, and there was no apparent difference in speed drop during peak hours and off-peak hours. The bad weather condition did not show a significant change of average speed, which meant weather was not a likely contributing variable in this study.

**TABLE 15 Significance Test for the Average Speed Change in Scenario Two**

Condition ID	Time of day	VMS Status	Weather	Speed Change	Z-statistic	p-value	Significance
6	Off Peak	Activated	None	-1.07	-3.637	$2.76e^{-4}$	***
7	AM Peak	Activated	None	-1.37	-2.805	$5.03e^{-3}$	***
8	PM Peak	Activated	None	-1.27	-3.024	$2.50e^{-3}$	***
9	Off Peak	Deactivated	None	-0.04	-0.116	0.908	
10	Off Peak	Activated	Bad	-0.75	-0.796	0.426	

**Notes: Significance codes: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, 0.1 ‘ ’ 1**

In summary, speed changes for eight different combinations of conditions were estimated (two of the ten conditions were the same). In terms of the regression outcomes, average speed change was more sensitive to the VMS status variables than the other two groups of variables. No significant change of speed was detected under all possible conditions when a VMS was deactivated. On the contrary, the models captured the significant change of speed when a VMS was activated at each time of day when the weather was normal. However, the magnitude of number was small (all less than 2.0 mph). Weather variables did not have an impact on the outcomes.

*c. Goodness of Fit for Analysis model*

Typically, for normal linear models, the adequacy of model is tested by checking the residuals. Residuals are considered effective tools for the goodness of fit for models, and they are supposed to be independent and have an approximate normal distribution with the mean zero and constant variance. (Dobson and Barnett, 2008)

Models for scenario one and scenario two were complementary. The general model was

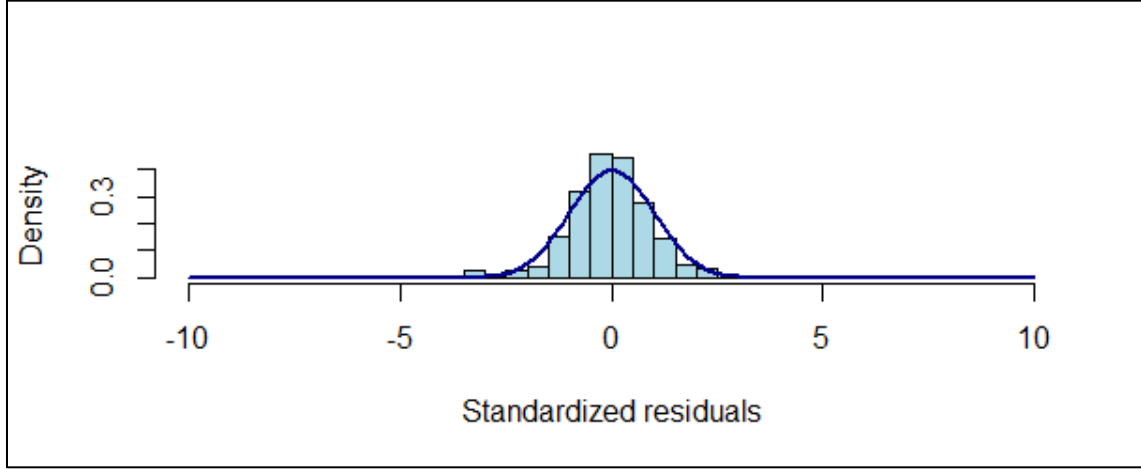
$$E(Y_i) = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}; \quad Y_i \sim N(\mu_i, \sigma^2). \quad (4.14)$$

The approximate standardized residuals are defined as equation (4.15), where  $\hat{\sigma}$  is the estimate of  $\sigma$  (Dobson and Barnett, 2008)

$$r_i = (y_i - \hat{\mu}_i) / \hat{\sigma}, \quad (4.15)$$

The distribution of  $r_i$  is displayed in FIGURE 11, and a normal distribution with the mean and standard deviation of  $r_i$  was added. Approximately, the distribution of  $r_i$  was Normal. In addition, the percentage of  $r_i$  that greater than 1.96 is 1.34%, and the

percentage of  $r_i$  that less than -1.96 is 2.54%. The frequencies of unusual values were not prevalent in this model.



**FIGURE 11 Histogram of Standardized Residuals**

#### 4.3.3 Analysis of Covariance

The analysis of difference scores stated in the last section was performed under the assumption that the average speeds before and after the activation or deactivation of VMS were the same. However, if the trend of observed initial speed ( $x_{bi}$ ) and impact speed ( $x_{ai}$ ) were not consistent under multiple scenarios, the results discussed in the last section would be not defensible. In other words, if the initial speed significantly affected the change in speed on analysis that assumes a constant change could be misleading. Thus, the analysis of covariance was performed in this section to test the impact of initial speeds on the outcomes.

The mathematical model for the analysis of covariance is

$$E(Y_i) = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}; \quad Y_i \sim N(\mu_i, \sigma^2) \quad (4.16)$$

$$\mu_i = \beta_0 + \sum_{j=1}^{3-1} \beta_{1j} T_{1j} + \sum_{k=1}^{2-1} \beta_{2k} S_{2k} + \sum_{m=1}^{2-1} \beta_{3m} W_{3m} + \beta_4 X_{bi} \quad (4.17)$$

Where

$Y_i$ = average speed after the change of VMS status;

$T_{1j}$ = time of day factor,  $j=1, 2, 3$ ;

$S_{2k}$ = VMS status factor  $k=1, 2$ ;

$W_{3m}$  = weather factor.  $m=1, 2$  (“good weather”, “bad weather”);

$X_{bi}$ = average speed before the change of VMS status; and

$\beta$ = coefficients.

**TABLE 16 Comparison of Regression Results for Difference Scores Model and Analysis of Covariance Model (Scenario One Example)**

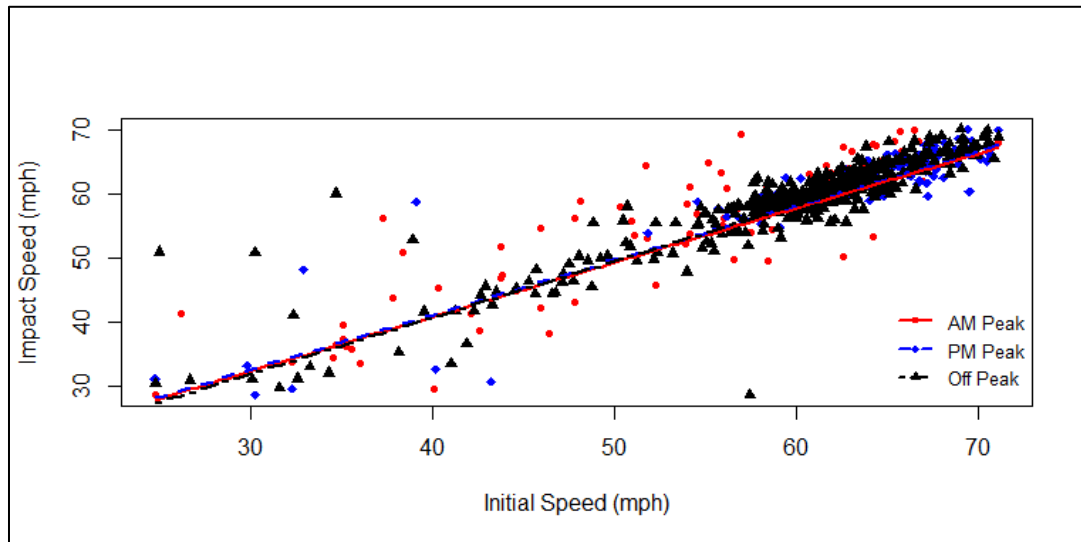
	Variable	Coefficient	Std. Error	t value	Pr (> t )	Significance
Difference Scores Model	(Intercept)	-0.037	0.320	-0.116	0.907	
	$t_{ai}$	-0.294	0.517	-0.568	0.570	
	$t_{pi}$	-0.195	0.479	-0.407	0.684	
	$S_{ai}$	-1.034	0.385	-2.685	0.007	**
	$W_{Bi}$	0.326	0.933	0.350	0.727	
Analysis of Covariance Model	(Intercept)	3.917	1.085	3.611	$3.28e^{-4}$	***
	$t_{ai}$	-0.585	0.518	-1.130	0.259	
	$t_{pi}$	-0.038	0.476	-0.079	0.937	
	$S_{ai}$	-0.944	0.382	-2.472	0.014	*
	$W_{Bi}$	0.132	0.925	0.142	0.887	
	$X_{bi}$	0.934	0.017	54.043	$<2.00e^{-16}$	***

Notes: Significance codes: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, 0.1 ‘ ’ 1

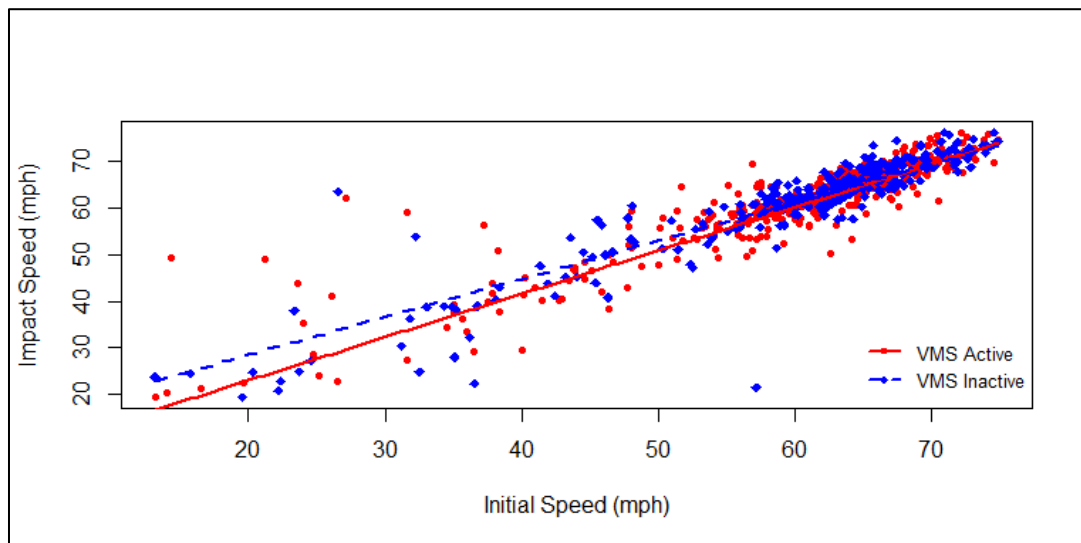
The regression was performed by R, and the results comparison was conducted between the difference scores model and the analysis of covariance model (TABLE 16). By looking at the statistics of the variable  $X_{bi}$ , we find that the initial speed was highly correlated with the impact speed (with a p-value smaller than  $2.00e^{-16}$ ). In addition, the coefficient (0.934) was close to 1.00, which indicated that there was roughly no difference between those two speeds. One the other hand, the patterns for the coefficients of variables are similar for both models. From the standpoint of practical significance, the comparison results indicated that the difference scores model was sufficient to be utilized for estimating the change of speed in various scenarios.

In order to further investigate the interactions among the initial speed variable and the other categorical explanatory variables, we can plot the relationship between initial speed and the impact speed respectively for each category. FIGURE 12 through FIGURE 14

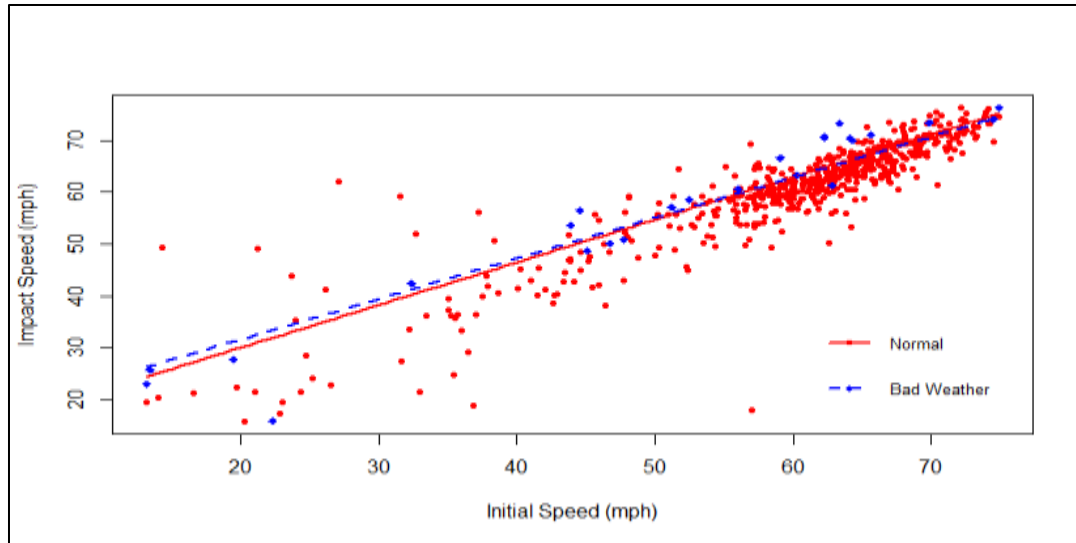
describe the relationships under different conditions. Speed curves were estimated by fitting the simple linear models, with the impact speed as the response and the initial speed as the explanatory variable.



**FIGURE 12 Speed Curves for Different Times of Day**



**FIGURE 13 Speed Curves for Different VMS Statuses**



**FIGURE 14 Speed Change for Different Weather Conditions**

The curves are roughly parallel suggesting that no apparent interactions between the initial speed and other interest variables were detected in terms of the plots. Thus, it was appropriate to use the difference scores models. It is notable that, in FIGURE 13, we found that the trendline for VMS active was lower than for VMS inactive, which was consistent with the outcomes from the analysis of difference scores. It indicated that the decrease of speed was higher when a VMS was activated.

In summary, the analysis of covariance showed that the difference scores model in the previous section was appropriate to be used for estimating the change of speed. From the practical significance standpoint, no significant change of model outcomes was observed when the two models were compared.

#### **4.4 Discussion**

The impact of VMS on short-time fluctuation in speed was analyzed in this study, and the results revealed that speed changes were not substantial. The analysis was conducted with two models: difference scores model, and analysis of covariance model. The comparison of the two models indicated that the interactions among the initial speed and other categorical explanatory variables were not significant. Thus, difference scores model was used to estimate the changes of speed in multiple scenarios, and to test the significance of the changes.

According to model results, there was no significant change of speed when VMS was deactivated in both peak hours and off-peak hours under all weather conditions. Speed decreases were observed in the VMS impact regions when the signs were activated, and the reductions were statistically significant. However, the magnitude of decrease was trivial. In addition, the difference of average speed change between off-peak hours and peak hours were not obvious. Moreover, model results discovered that weather conditions were not potential contributing factors to affect the speed under the impact of VMS. In summary, drivers were likely to decelerate when noticed sign messages were displaying ahead. However, the decrease of speed was not expected to be substantial, normally within 2.0 mph. The average reduction of speed estimated in this study was lower than the decrease from the previous studies (Haghani et al., 2013; Erke et al., 2007; Harder et al., 2003). One of the possible reasons was that the samples with significant decrease of speed, which were potentially related to downstream shockwaves, were identified, and removed from the model in this study. Those samples were not included because VMS was not a single factor that affected the change of speed, if a shockwave reached the VMS impact region when a message was displaying. However, there were some “outliers” included in the models. The “outliers” were cases with speed decrease more than 20 mph. Those substantial speed decreases might be attributed to many reasons. First, the speeds were average speed extracted from loop detectors rather than measured for individual vehicles. Detectors problems were possible reasons for the variation of speed. In addition, some other downstream-unknown events, such as crashes, which were not posted in the VMS messages, could also cause the decrease of speed. The models in this study included those “outliers”, and results revealed that they did not affect the model outcomes significantly.



## **Chapter 5: Case-Control Study**

Although no apparent change of average speed in the VMS impact regions was observed when the signs were activated and deactivated according to the Chapter 4, there is a possibility that drivers are distracted by the VMS messages without slowing down. Distraction is a potential risk factor for crashes. In the study of Bakiri et al. (Bakiri et al., 2013), 8% of injurious road crashes were associated with distracting events inside the vehicles, like making phone calls and picking up objects; and events outside the vehicles, like looking at certain non-recurring events. However, Bakiri did not mention the overall percentage of distractions that result in, and did not result in crashes, so we do not know the supposed “distraction object” is a risk factor or not. In this study, VMS messages are possible outside distraction factors, which could be associated with higher probability of crash occurrence.

Previous studies have tended to overlook the impact of VMS on crash occurrence. In part, crash data collected from the police reports was not precise enough to match with VMS data. In this chapter, a case-control study was performed, and odds ratios were derived to determine if VMS activation is overrepresented in crashes. Crash times were estimated rather than using the reported time directly.

### **5.1 Data Preparation**

#### **5.1.1 VMS Logs Data**

Five VMS logs along I-94 from downtown Minneapolis to downtown St. Paul were utilized for this study (See TABLE 1). For each VMS site, the log data were collected from January 2006 to December 2012. The critical information needed was the VMS messages, together with their times of deployment and clearance.

The raw data from VMS logs, which were requested from MnDOT, did not provide the ending time directly for each message. However, records for “Sign DEPLOYED” and “Sign CLEARED” were considered two types of events, and were recorded separately. Hence, if the a sign was activated and a message was displayed, the event time of this record was the starting time of the message, and the event time for the next record which describes the sign clearance was supposed to be the ending time. Ideally, Records for

“Sign DEPLOYED” and “Sign CLEARED” were paired and the time intervals between each other were reasonable. Validation for the processed VMS data was conducted by checking the duration of the messages based on the developed starting time and ending time. Records were removed from the analysis if the duration was not practical.

In total, 7849 usable messages with the time of activation and deactivation, from five VMS logs in 7 years, were identified.

### **5.1.2 Crash Data**

Crash data were collected from MnCMAT in the VMS impact regions from January 2006 to December 2012. One of the major advantages of MnCMAT is the adequate crash-related information. Typically, one record in MnCMAT includes information for the crash such as occurrence time, location, and severity; information on drivers such as age and gender; information on vehicles such as vehicle type and direction; and information on other conditions such as lighting and weather conditions.

Times and locations of crashes were the primary attributes utilized in this study. Crashes times were used to determine whether they occurred when the VMS was active. Instead of working with the published times, estimated crash times were used in the case-control study due to the inaccuracy of the police reports. The methodology for crash time estimation was stated in Chapter 3.4.1, and this method was implemented for all the crashes collected. Locations of the crashes were developed from the attribute named “True\_Miles” in MnCMAT database, and this information was used to determine whether crashes occurred in the VMS impact regions. However, the reported locations were not precise in a few of the crash records because drivers might move their vehicles to the shoulders or other safe places nearby after the crashes. Those exceptions could be noticed by plotting the traffic data (like occupancy) measured from detectors close to the VMS impact regions. If the mile markers indicated crashes were in the VMS impact regions but no fluctuation in traffic was identified by loop detectors, the crashes were not included.

In total, 183 crash records were collected for the VMS impact regions in 7 years. All of the crashes were considered “cases”. Microsoft Excel Macro was applied to link the crash records and VMS logs. In summary, for each crash record, there were only two options

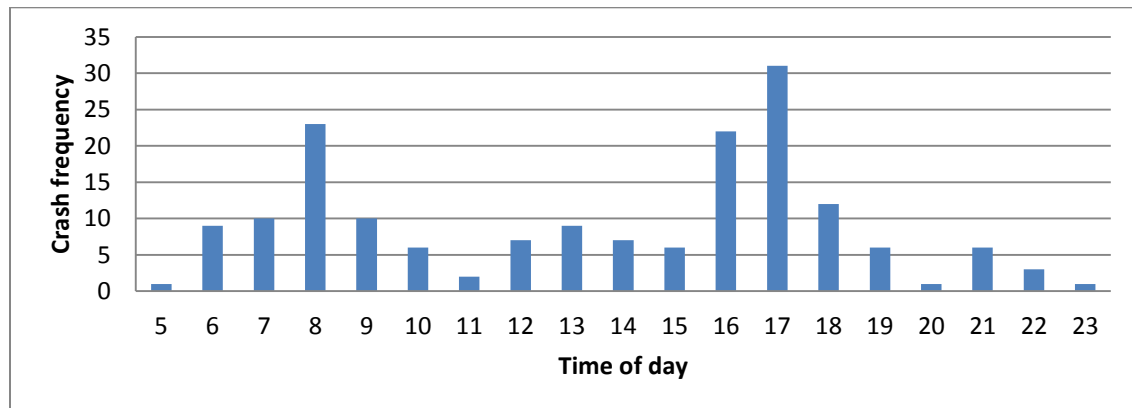
for VMS status: active or blank. After running the Macro codes, 25 crashes (13.7%) occurred when the signs were active.

### 5.1.3 Non-crash Controls

The control group was a cluster of non-crash events that were similar to “cases” on certain specific characters. In this study, the exposures of controls that need to be matched with cases were time of the crashes, and traffic conditions, both of which were possible relevant factors to the crashes. Four controls were selected for each case since there was a general argument that no significant statistical power increasing beyond the 4:1 control-case ratio. (Zheng, 2010; Schlesselman, 1982)

#### *a. Controls Matching Times of the Cases*

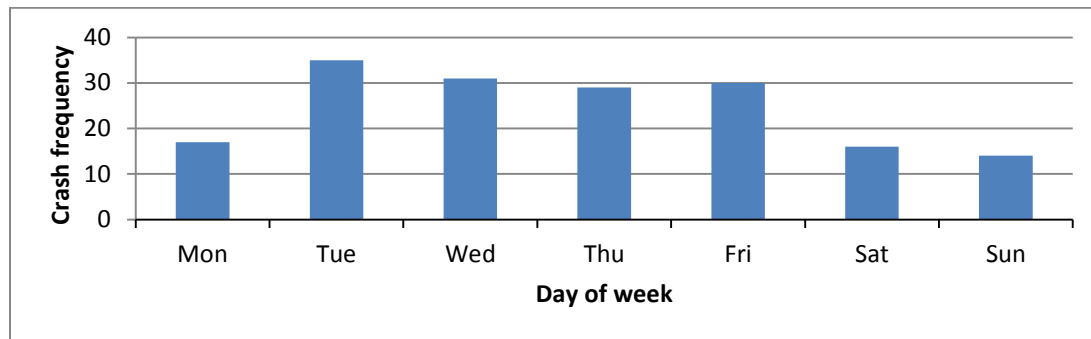
The frequencies of crashes vary depending on the time of day, and day of week (FIGURE 15 and FIGURE 16). Generally, crashes were likely to occur during the peak hours on Tuesday to Friday in the selected VMS impact regions. Hence, the control of event time is critical. The starting time of the controls was selected to be at the same time of the day and the same day of week as the starting time of the cases, but was collected in adjacent four weeks.



**FIGURE 15 Crash Frequencies by Time of Day (All 183 Crashes Collected from five VMS Sites)**

For example, assuming that the starting time of a case was 9:00 AM on June 15th 2012 (Friday). The starting time of controls would be 9:00 AM on June 8th 2012 (last Friday), 9:00 AM on June 1st 2012 (one week before last Friday), 9:00 AM on June 22nd 2012

(next Friday), and 9:00 AM on June 29th 2012 (one week after next Friday). However, to avoid the overlapping of crashes, those controls were valid only if there was no crash occurred at least 10 minutes before their starting times.



**FIGURE 16 Crash Frequencies by Day of Week (All 183 Crashes Collected from five VMS Sites)**

*b. Controls Matching Traffic Conditions Prior to the Cases*

Ideally, traffic patterns are similar at a specific time of week. Traffic conditions could be automatically matched if the time of the cases and controls are fixed at the same point. However, there was an argument that traffic conditions prior to the crashes were not consistent with the normal conditions. A higher probability of crashes was associated with stronger fluctuation in downstream traffic. Standard deviation of downstream speed and occupancy were considered dangerous factors that were associated with upstream crashes (Chapter 2.2). Therefore, traffic conditions prior to the cases and controls need to be further investigated in order to guarantee the quality of the controls.

Over manually-constraining the data was not recommended in this study. However, controls with some “special unknown events” were removed from the database. Those events were unknown due to the limitation on information, but they could cause the traffic conditions to differ between the controls and cases. Fortunately, those conditions could be noticed by comparing the plots of occupancy measured by loop detectors before the starting times of the cases and controls, although the types of the events were unknown. Controls were removed from the database when “special unknown events” were detected. Moreover, if all the controls of a case were deleted, the particular case was removed as well.

In sum, 183 cases and 732 controls (two of the controls were overlapped with cases) were usable in this study. Based on the obtainability of the traffic data and the rules of selecting controls stated above, 172 cases and 631 controls were available for the logistic regression model with various traffic metrics variables.

## 5.2 Methodology and Model Construction

The case-control study was used to assess the risk of crashes when VMS was active. Cases were the crashes that occurred in the VMS impact regions, and controls were non-crash events, which had similar conditions to cases. First,  $2 \times 2$  contingency tables (Agresti, 1990) were constructed to specify the relationship between VMS status and events. Then the corresponding unadjusted odds ratios (Agresti, 1990) were derived; and the hypothesis that the deployment of VMS had no impact on crash occurrence was tested. Finally, the adjusted odds ratios were estimated by performing a logistic regression; and the significance of the results was investigated.

### 5.2.1 $2 \times 2$ Contingency Table

TABLE 17 describes the distribution of two categorical response variables, both of which have two categories. Let  $V$  denotes the condition that VMS is active,  $\bar{V}$  denotes the condition that VMS is blank,  $C$  denotes there is crash observed, and  $\bar{C}$  denotes there is no crash observed. Cell frequency is defined as  $n_{ij}$ . For example,  $n_{11}$  means that there are  $n_{11}$  crashes occurred when VMS was active among all the counts.

Instead of using cell frequencies, the contingency table can also be displayed with joint distribution of the events and VMS status. Generally, this joint distribution is denoted by  $\pi_{ij}$ , and the marginal distributions are derived by adding the row and column joint probabilities, which are identified by  $\pi_{i+}$  and  $\pi_{+j}$ . (Agresti, 1990) Assuming that the VMS status and events are independent, then

$$\pi_{ij} = \pi_{i+} \pi_{+j} \quad \text{for } i = 1, 2 \text{ and } j = 1, 2 \text{ (Agresti, 1990)}$$

The information in TABLE 17 could be converted into the joint and marginal probabilities in TABLE 18.

**TABLE 17 Cross-Classification of VMS Statuses and Crashes**

VMS status	Events		Total
	Crashes ( $C$ )	Non-crash ( $\bar{C}$ )	
VMS is active ( $V$ )	$n_{11}$	$n_{12}$	$n_{11} + n_{12}$
VMS is blank ( $\bar{V}$ )	$n_{21}$	$n_{22}$	$n_{21} + n_{22}$
<b>Total</b>	$n_{11} + n_{21}$	$n_{12} + n_{22}$	<b><math>n</math></b>

**TABLE 18 Joint and Marginal Probabilities of VMS Statuses and Crashes**

VMS status	Events		Total
	Crashes ( $C$ )	Non-crash ( $\bar{C}$ )	
VMS is active ( $V$ )	$\pi_{11} = n_{11}/n$	$\pi_{12} = n_{12}/n$	$\pi_{1+} = \pi_{11} + \pi_{12}$
VMS is blank ( $\bar{V}$ )	$\pi_{21} = n_{21}/n$	$\pi_{22} = n_{22}/n$	$\pi_{2+} = \pi_{21} + \pi_{22}$
<b>Total</b>	$\pi_{+1} = \pi_{11} + \pi_{21}$	$\pi_{+2} = \pi_{12} + \pi_{22}$	<b>1.0</b>

### 5.2.2 Odds Ratio

The odds ratio was used in this study to quantify the strength of association between the activation or deactivation of VMS and the crashes in a given population. According to Agresti (1990), for a  $2 \times 2$  contingency table with cell probability of  $\pi_{ij}$ , the odds ratio can be derived as

$$\theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} . \quad (5.1)$$

For contingency table with cell counts of  $n_{ij}$ , the odd ratio can also be estimated as (Agresti, 1990)

$$\hat{\theta} = \frac{n_{11} n_{22}}{n_{12} n_{21}} . \quad (5.2)$$

The impact of VMS on crash occurrence can be simplified by interpreting the odds ratios. For example, if  $\theta = 2.0$ , it means that the odds of crash are twice that when a VMS is active comparing to a VMS is blank. However, it only indicates that crashes are more likely to occur when VMS is activated. We cannot say that the probability of crash is twice greater than in non-VMS conditions. Nonetheless, the deployment of VMS is a potential risk factor to crashes if the odds ratio is greater than 1.0, and a higher odds ratio shows a stronger association.

The hypothesis, which the odds ratio is not significantly different from 1.0, is required to be tested. According to the previous studies, the logarithm transformation of the estimated odds ratio,  $\log \hat{\theta}$ , have an asymptotic normal distribution around  $\log(\theta)$ , but it converges much faster than  $\hat{\theta}$  (Agresti, 1990). The estimated standard error for  $\log \hat{\theta}$  is

$$\hat{\sigma}(\log \hat{\theta}) = \left( \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right)^{\frac{1}{2}} . \quad (5.3)$$

With the logarithm of estimated odds ratio  $\log \hat{\theta}$ , and the estimated standard error of  $\log \hat{\theta}$ , a hypothesis test can be performed with the null hypothesis that the odds ratio is equal to 1.0, and the alternative hypothesis that the odds ratio is not equal to 1.0. If taking logarithm for both sides, those hypotheses can be transformed into

$$\begin{aligned} H_0: \quad & \log \hat{\theta} = \log(1.0) = 0 \\ H_A: \quad & \log \hat{\theta} \neq \log(1.0) = 0 . \end{aligned} \quad (5.4)$$

The test statistic,  $Z$  is

$$Z = \frac{\log \hat{\theta} - 0}{\hat{\sigma}(\log \hat{\theta})} = \frac{\log \hat{\theta}}{\hat{\sigma}(\log \hat{\theta})} . \quad (5.5)$$

When  $n$  is great enough, the test statistic is approximately a standard normal

$$Z \sim N(0, 1).$$

Then the critical Z-value could be derived from the standard normal table, and the significance of the estimated odds ratio can be tested. For example, the critical Z-value equals to  $\pm 1.96$  at the significance level of  $\alpha = 0.05$ . If the calculated Z-value based on the  $2 \times 2$  contingency table is smaller than 1.96 and greater than -1.96, for instance 1.00, there is no strong evidence to reject the null hypothesis. This indicates that the odds ratio is not significantly different from 1.0. In this case, no apparent association between the deployment of VMS and crash occurrence is detected. On the contrary, a higher absolute Z-value indicates a probable association.

### 5.2.3 Logistic Regression Model

In this study, the goal of using logistic regression model is to recognize the relationship between the crash occurrence and a set of explanatory variables, and to derive the adjusted odds ratio for the VMS while controlling for the impacts of other predictors.

Suppose the response variable,  $Y_i$ , has a bernoulli distribution

$$Y_i \sim B(1, \pi_i), \quad (5.6)$$

where  $\pi_i$  is the probability of being a case. The multiple logistic regression model is expressed by the Equation (5.7) (Hosmer and Lemeshow, 2000)

$$\begin{aligned} \text{logit}(\pi_i) &= g(x) \\ &= \beta_0 + \sum \beta_{1i} x_{VMS} + \sum \beta_{2i} x_{Traffic\ Conditions} \\ &\quad + \sum \beta_{3i} x_{Weather} + \sum \beta_{4i} x_{Crash\ Type} + \sum \beta_{5i} x_{Sites} \end{aligned} \quad (5.7)$$

Where  $x_i$  are explanatory variables,  $\beta_i$  are the corresponding coefficients, and

$$\pi_i = \frac{e^{g(x)}}{1 + e^{g(x)}}. \quad (5.8)$$

With the logit link function and the probability of the response variable, R is capable to estimate the parameters of the model, and to provide useful information for testing the



significance of the coefficients. The objective of using logistic regression is to investigate the association of VMS and crash occurrence rather than predicting crash. Hence, we kept the VMS variables in certain models for estimating the adjusted odds ratios even if their coefficients were not significant.

Since VMS is considered a dichotomous independent variable with the code 0 and 1, the adjusted odds ratio can be estimated with its regression coefficient (Hosmer and Lemeshow, 2000)

$$\hat{\theta} = e^{\beta_{VMS}} . \quad (5.9)$$

Where,

$\hat{\theta}$  = estimated odds ratio of VMS, and

$\beta_{VMS}$  = coefficient of VMS variable.

### 5.3 The Deployment of VMS and Crash Occurrence

#### 5.3.1 Unadjusted Odds Ratio

For this study, 183 cases from the five selected VMS impact regions were available. Four controls were collected corresponding to one case. With the crash data and VMS logs, VMS statuses were linked to both types of events. Without any causal factor adjusted among cases and controls, the 2×2 contingency table is developed as TABLE 19. The total number of controls is slightly smaller than the four times of the total number of cases. This is because that some controls were overlapped with other cases.

**TABLE 19 Distribution of Crashes and Non-crash Controls**

VMS status	Events		Total
	Crashes ( $C$ )	Non-crash ( $\bar{C}$ )	
VMS is active ( $V$ )	25	82	107
VMS is blank ( $\bar{V}$ )	158	648	806
<b>Total</b>	183	730	913

Recall the Equation (5.2), the sample odds ratios is derived as

$$\hat{\theta}_0 = \frac{n_{11} n_{22}}{n_{12} n_{21}} = \frac{25 \times 648}{158 \times 82} = 1.250 \quad (5.10)$$

This result reveals that the odds of crash are 1.250 times when the VMS is active than is blank. However, it does not indicate that the probability of crash is 1.250 times higher with the VMS was active due to the inequity of odds ratio and relative risk. The relationship between them is (Agresti, 1990)

$$\text{odds ratio} = \text{relative risk} \left( \frac{1 - \pi_2}{1 - \pi_1} \right) . \quad (5.11)$$

Where,

$\pi_1$  = the probability of crash under the condition that the VMS is active; and

$\pi_2$  = the probability of crash under the condition that the VMS is blank.

However, the probability of crashes cannot be estimated given whether crashes occurred when VMS was active or inactive. Thus, the relative risk was unknown. Nonetheless, with an estimated odds ratio that was greater than 1.0, it is probable that drivers would experience higher risk when driving in the VMS impact regions if the signs were activated.

Although the estimated odds ratio is greater than 1.0, this does not mean the estimator is significant since it is still close to 1.0. According to Equation (5.5), the test statistic  $Z_0$  is

$$Z_0 = \frac{\log \hat{\theta}_0}{\hat{\sigma}(\log \hat{\theta}_0)} = \frac{\log \hat{\theta}_0}{\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}} = 0.912 . \quad (5.12)$$

At the significant level of  $\alpha = 0.05$ , the critical Z-value is  $\pm 1.96$ . However,  $Z_0 = 0.912$  shows that the estimated odds ratio is not significantly different from 1.0. This indicated no strong evidence for rejecting that the probability of crashes when VMS was active was the same with the probability when VMS was blank. This fact was comparable if each VMS impact region was analyzed separately (See TABLE 20). None of the selected regions shows significant associations between the deployment of the VMS and the crash occurrence.

**TABLE 20 Odds Ratios and Hypothesis Test Results for Individual VMS Impact Regions**

VMS ID	Odds Ratio	Z-Value	Significance at $\alpha = 0.05$
V94W09	1.06	0.145	Not Significant
V95W07	2.34	1.592	Not Significant
V94W05	0.65	-0.394	Not Significant
V94E11	2.11	0.584	Not Significant
V95E08	1.15	0.293	Not Significant

However, the major limitation of using the unadjusted odds ratio is that other causal factors such as traffic prior to the events and weather conditions were not considered in the analysis. Typically, freeway crashes are related to downstream shockwaves, where a speed difference is generated at downstream traffic. Since the unadjusted controls were selected randomly based on the starting time of the cases, it is possible that the traffic conditions for some of the controls were free-flow traffic when the probability for crashes was low. This fact could potentially affect the value of odds ratio, in other words, underestimate or overestimate the impact of VMS on crash occurrence. In addition, adverse weather conditions are other possible risk factors to the crashes, which were not controlled in the analysis.

### **5.3.2 Adjusted Odds Ratio**

#### *a. Variable Selection*

The selection of variables is generally the first step for performing logistic regression. Variables included in the models were supposed to be associated with the crash occurrence, but not highly correlated with the VMS factor, which was the interest variable. The critical causal factors used in this study included various traffic metrics

such as downstream standard deviation of speed and occupancy, and adverse weather conditions such as rainfall and snowfall.

Based on the availability of the data in this study, seven traffic metrics, derived at sites downstream from crashes, could be used as explanatory variables for the logistic regression model (See TABLE 6). However, it is neither efficient nor effective to regress all of them in the model without knowing their properties individually. This is because some of them are highly correlated with each other, for example, the coefficient of variation (COV) in speed was derived as the ratio of standard deviation of speed and average speed. In addition, certain variables probably have limited association with crash occurrence, which need to be excluded.

According to Hosmer (2000), the selection of variables for a logistic model should start with a univariable analysis of each variable, and a variable is included in the model if its univariable test has a p-value smaller than 0.25. The univariable logistic regression results are displayed in the TABLE 21. Other than downstream flow, all the p-values of the metrics are smaller than 0.25, and five of six are smaller than 0.05. Hence, the variable of downstream traffic flow was excluded from the model. Moreover, as stated above, COV in speed is the ratio of average speed and standard deviation of speed. It is not appropriate to include highly correlated explanatory variables in the same model. In this study, the average speed was removed, since the other two metrics were argued to be preferable crash predictors than average states (Zheng et al., 2010).

**TABLE 21 Univariable Logistic Regression Results for Selected Traffic Metrics**

Traffic Metrics ID	Coefficient	Std. Error	Z-value	p-value	Significance
AvgS_down	-0.018	0.005	-3.573	0.000	***
AvgOcc_down	0.024	0.008	3.046	0.002	**
StdS_down	0.069	0.033	2.090	0.037	*
StdO_down	0.053	0.018	2.985	0.003	**
COV_S_down	1.542	0.530	2.911	0.004	**
MAXDiff_S_down	0.016	0.013	1.217	0.224	
Flow_down	-0.270	0.236	-1.146	0.252	

**Notes: Significance codes: 0 '\*\*\*', 0.001 '\*\*', 0.01 '\*', 0.05 '.', 0.1 ' ' 1**

Other than traffic metrics, variables of weather conditions, crash diagrams, and VMS sites were also considered in the regression models. These variables were discrete,

nominal scale variables, and the numbers of the variable had no numeric significance. However, R was capable to generate an appropriate collection of dummy variables. In addition, instead of aggregating all types of VMS messages, the logistic regression model can estimate the coefficient of each type of VMS messages separately. This was used to test whether the impact of VMS on traffic depends on the types of messages displayed on the signs.

**TABLE 22 Possible Explanatory Variables for Logistic Model**

	Variable ID	Description
Traffic metrics	MAXDiff_S_down	Downstream maximum speed difference of adjacent lanes
	COV_S_down	Downstream coefficient of variation in speed
	AvgOcc_down	Downstream average occupancy
	StdS_down	Downstream standard deviation of speed
	StdO_down	Downstream standard deviation of occupancy
Precipitation types	Rain	Observed rainfall
	Snow	Observed snowfall
	Frozen	Observed frozen on road (thin ice)
	None	No adverse weather condition observed
Crash diagrams	Diag <sub><i>i</i></sub>	Crash type <i>i</i>
VMS sites	VMSID <sub><i>i</i></sub>	VMS Site <i>i</i>
VMS messages	VMS1	Danger/Warning messages
	VMS2	Informative/Common messages
	VMS3	Regulatory/Non-Traffic messages
	Blank	No message

In summary, the possible explanatory variables for logistic model are listed in TABLE 22, which contains 5 traffic metrics variables, 4 weather variables (3 dummy variables), 8 crash diagram variables (7 dummy variables), 4 VMS site variables (3 dummy variables), and 4 VMS message variables (3 dummy variables). The association between the types of VMS messages and crash occurrence was the primary concern in this study. In reality, no VMS message of type three was observed among cases and controls. Hence, there were actually 3 VMS message variables (2 dummy variables).

*b. Logistic Regression and Hypothesis Tests*

Let  $Y$  denote the crash outcome, and  $x_i$  denote the explanatory variables in TABLE 22. Assuming that the response variable has a bernoulli distribution (Equation 5.6), the multiple logistic regression model is defined as

$$\begin{aligned} \text{logit}(\pi_i) &= g(x) \\ &= \beta_0 + \sum_{j=1}^5 \beta_{1j}x_{1j} + \sum_{k=1}^{4-1} \beta_{2k}x_{2k} + \sum_{m=1}^{8-1} \beta_{3m}x_{3m} + \sum_{n=1}^{4-1} \beta_{4n}x_{4n} + \sum_{p=1}^{3-1} \beta_{5p}x_{5p} , \end{aligned} \quad (5.13)$$

Where

$\pi_i$ = expected probability of crash;

$x_{1j}$ = variable of traffic metrics  $j, j=1, 2, 3, 4, 5$ ;

$x_{2k}$ = variable of precipitation types  $k, k=1, 2, 3, 4$ ;

$x_{3m}$ = variable of crash diagram  $m, m=1, 2, \dots, 8$ ;

$x_{4n}$ = variable of VMS sites  $n, n=1, 2, 3, 4$ ;

$x_{5p}$ = variable of VMS messages  $p, p=1, 2, 3$ ; and

$\beta$ = coefficients.

The regression results are summarized in the TABLE 23. At the significance level  $\alpha=0.05$ , the outcomes revealed that the coefficients of crash type variables, VMS sites variable, and VMS messages variables were not statistically significant, which indicated that the change of those variables would not affect the variation of crash probability. Conversely, certain traffic metrics variables and weather variables showed significance as expected. Variables of downstream occupancy, standard deviation of speed, and COV in speed were found to be significantly related to the probability of crash occurrence as well as rainfall and snowfall.

According to the regression results stated in TABLE 23, a new logistic regression was performed with the explanatory variables whose coefficients were significant or close to significant (such as the downstream standard deviation of occupancy, and the maximum difference speed whose p-values were close to 0.1). The new model was called “small

model”, and the previous model that included all the selected variables was called “big model”. In the small model, all the assumed non-relevant variables were removed including the VMS messages types, which were the variables of interested. The regression results of the small model are displayed in TABLE 24. The influence of the removed variables can be tested by the comparison of the big model and small model.

**TABLE 23 Regression Results for All the Selected Explanatory Variables (“Big Model”)**

Variable ID	Coefficient	Std. Error	z value	Pr(> z )	Significance
(Intercept)	-3.714	0.592	-6.272	$3.57e^{-10}$	***
AvgOcc_down	0.061	0.022	2.751	$5.94e^{-3}$	**
StdS_down	0.253	0.074	3.432	$6.00e^{-4}$	***
StdO_down	0.082	0.053	1.562	0.118	
COV_S_down	-4.476	2.138	-2.094	0.0363	*
MAXDiff_S_down	-0.039	0.025	-1.573	0.116	
Rain	1.264	0.392	3.226	$1.26e^{-3}$	**
Frozen	-0.283	1.135	-0.25	0.803	
Snow	3.825	0.814	4.697	$2.64e^{-6}$	***
diag2	-0.005	0.249	-0.019	0.985	
diag4	-0.215	0.510	-0.421	0.674	
diag5	-0.409	0.994	-0.411	0.681	
diag7	0.013	0.524	0.025	0.980	
diag8	0.140	1.197	0.117	0.907	
diag90	0.001	0.378	0.003	0.998	
diag98	0.816	1.152	0.708	0.479	
V94W05	0.459	0.318	1.442	0.149	
V94W07	0.147	0.304	0.483	0.629	
V94W09	-0.332	0.286	-1.161	0.246	
Message Type1	0.222	0.298	0.744	0.457	
Message Type2	0.249	0.592	0.421	0.674	

**TABLE 24 Regression Results for “Small Model”**

Variable ID	Coefficient	Std. Error	z value	Pr(> z )	Significance
(Intercept)	-3.467	0.534	-6.488	$8.67e^{-11}$	***
AvgOcc_down	0.064	0.021	3.051	0.002	**
StdS_down	0.246	0.073	3.352	0.001	***
StdO_down	0.055	0.049	1.123	0.262	
COV_S_down	-4.387	2.050	-2.140	0.032	*
MAXDiff_S_down	-0.047	0.023	-2.008	0.045	*
Rain	1.240	0.386	3.209	0.001	**
Snow	3.863	0.799	4.835	$1.33e^{-6}$	***

Notes: Significance codes: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, 0.1 ‘ ’ 1

By roughly comparing the coefficients and z-values of the explanatory variables in big model and small model, no apparent change of magnitude was detected. This indicated that the removed variables did not significantly affect the outcomes. If we used stepwise model selection (Hosmer and Lemeshow, 2000), and added the VMS message variables back to the small model, the results turned out to be still no significant difference. (See TABLE 25). For clarification, this model with VMS messages variables included was called “medium model”.

**TABLE 25 Regression Results for “Medium Model”**

Variable ID	Coefficient	Std. Error	z value	Pr(> z )	Significance
(Intercept)	-3.491	0.536	-6.519	$7.08e^{-11}$	***
AvgOcc_down	0.066	0.021	3.093	0.002	**
StdS_down	0.246	0.073	3.361	0.001	***
StdO_down	0.055	0.049	1.124	0.261	
COV_S_down	-4.547	2.062	-2.205	0.027	*
MAXDiff_S_down	-0.046	0.024	-1.931	0.053	.
Rain	1.263	0.388	3.253	0.001	**
Snow	3.866	0.801	4.826	$1.39e^{-6}$	***
Message Type1	0.154	0.293	0.525	0.599	
Message Type2	0.205	0.588	0.349	0.727	

**Notes: Significance codes: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, 0.1 ‘ ’ 1**

**TABLE 26 Logistic Regression Model Selection**

Models	Given statistics			Analysis of Variance (ANOVA)		
	Residual DF	Residual Dev	AIC	Difference DF	Difference Dev	p-value
Big model	782	766.19	808.19	11	6.51	0.837
Medium model	793	772.7	792.7	2	0.366	0.833
Small model	795	773.07	789.07			

A detailed comparison of the three models is displayed in TABLE 26. There was no significant difference of the residual deviance and Akaike information criterion (AIC) among the three models. In addition, according to the ANOVA, the data gave no evidence of statistically significant departure from the “smaller” models compared to the “bigger” models. Nonetheless, the objective of the logistic regression is to investigate the association between the VMS messages and crash occurrence rather than crash prediction.



The multiple regression results indicated that both of the coefficients of VMS messages types were not significantly different from zero.

Recall Equation (5.9), the adjusted odds ratios for VMS can be derived based on the medium model results, which were slightly better than big model. The odds ratios for the two types of VMS messages are

$$\text{VMS Message Type 1: } \hat{\theta}_1 = e^{0.15412} = 1.167,$$

$$\text{VMS Message Type 2: } \hat{\theta}_2 = e^{0.20540} = 1.228.$$

The 95% confidence interval (CI) for the estimated odds ratio is generally given by the expression (Hosmer and Lemeshow, 2000)

$$\exp[\hat{\beta}_i \pm z_{1-\alpha/2} \times \widehat{SE}(\hat{\beta}_i)] \quad (5.14)$$

Thus, the 95% CI for the estimated odds ratios are

$$\text{VMS Message Type 1: } \exp[0.15412 \pm 1.96 \times 0.29330] = (0.657, 2.073),$$

$$\text{VMS Message Type 2: } \exp[0.20540 \pm 1.96 \times 0.58781] = (0.388, 3.886).$$

There was still no significant association between the deployment of VMS and crash occurrence according to the adjusted odds ratios even the other possible causal factors were controlled. In addition, the change of different types of messages displayed on the signs showed no impact on the outcomes. The odds for crashes were slightly higher when a VMS was activated than a VMS was blank for each message types, but the difference was not significant. However, there is an argument that the sample size for the VMS “Message Type 2” is not sufficient (18 of 803 in total). Thus, it is necessary to investigate the association between the crash occurrence and the aggregated VMS message types. A new model was constructed which was similar with the medium model but the disaggregated VMS variables were replaced. The regression results are displayed in

TABLE 27. The aggregated VMS variable is not statistically significant. However, the patterns of other variables in TABLE 27 are consistent with the patterns in TABLE 25. These results indicated that the aggregated VMS message type was not likely a risk factor for crash.

**TABLE 27 Regression Results for “Medium Model” with Aggregated VMS Message Type**

Variable ID	Coefficient	Std. Error	z value	Pr(> z )	Significance
(Intercept)	-3.490	0.535	-6.521	7.00e <sup>-11</sup>	***
AvgOcc_down	0.065	0.021	3.093	0.002	**
StdS_down	0.246	0.073	3.362	0.001	***
StdO_down	0.055	0.049	1.128	0.259	
COV_S_down	-4.547	2.062	-2.206	0.027	*
MAXDiff_S_down	-0.046	0.024	-1.932	0.053	.
Rain	1.263	0.388	3.253	0.001	**
Snow	3.864	0.801	4.825	1.40e <sup>-6</sup>	***
Aggregated Message Type	0.163	0.270	0.605	0.545	

**Notes: Significance codes: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, 0.1 ‘ ’ 1**

The odds ratio for the aggregated VMS message types is

$$\text{Aggregated Message Types: } \hat{\theta}_a = e^{0.163} = 1.177.$$

Although the estimated odds ratio is greater than 1.0, it is not statistically significant. Based on the analysis of different models, there is no evidence implying that the VMS was a risk factor to freeway crashes.

*c. Goodness of Fit (GOF) of Logistic Regression Models*

Since the goal of the logistic regression is to investigate the relationship between the deployment of VMS and crash occurrence rather than crash prediction, it is not necessary to spend large efforts on pursuing the “best subsets” due to the insignificance of VMS variables. However, it is essential to guarantee the fit of the models. The Hosmer-Lemeshow tests (Hosmer and Lemeshow, 2000) were performed for GOF in this study.

Hosmer-Lemeshow tests are based on the comparison of observed outcomes and estimated probabilities with percentile-type grouping. Typically, the number of groups is defined as  $g = 10$ . The Hosmer-Lemeshow GOF statistics,  $\hat{C}$ , is obtained by deriving

Pearson chi-square statistics, and the corresponding p-value is usually used to test the null hypothesis that the model is actually correct (Hosmer and Lemeshow, 2000). In this study, the tests were performed with the number of defined groups ranging from 5 to 15. R package named “Resource Selection” was capable to provide Hosmer-Lemeshow GOF statistics, degree of freedom, and the corresponding p-values. The test results are displayed in TABLE 28. By looking at the p-values, none of them is smaller than 0.05. These results indicated that the null hypothesis that the models were corrected could not be rejected based on the significance level  $\alpha=0.05$

**TABLE 28 Hosmer-Lemeshow Tests Results for Medium Model**

Grouping Strategy		Medium Model		Small Model	
# Groups	DF	$\hat{C}$	p-value	$\hat{C}$	p-value
5	3	5.736	0.125	3.770	0.287
6	4	7.944	0.094	4.948	0.293
7	5	8.561	0.128	6.469	0.263
8	6	10.874	0.092	8.896	0.180
9	7	7.453	0.383	4.227	0.753
<b>10</b>	<b>8</b>	<b>9.498</b>	<b>0.302</b>	<b>9.036</b>	<b>0.339</b>
11	9	15.894	0.069	9.546	0.388
12	10	12.290	0.266	11.964	0.288
13	11	11.933	0.369	10.700	0.469
14	12	14.496	0.270	12.804	0.384
15	13	14.802	0.320	11.933	0.533

## 5.4 Discussion

A case-control study was performed to analyze the association between the deployment of VMS and crash occurrence. Based on the analysis of 2×2 contingency tables and logistic regression, the odds ratios were derived and their significance tested. The unadjusted odds ratio, which was estimated with the controls without the consideration of other causal factors, indicated that crashes were slightly more likely to happen in the VMS impact regions when the signs were activated. Nevertheless, this result was not significant after performing the hypothesis test at the significance level  $\alpha=0.05$ . The adjusted odds ratios were derived by fitting logistic models in order to control for traffic variables, weather variables, crash type variables, and VMS site variables. The regression

outcomes showed no evidence of the association between the deployment of VMS and crash occurrence after the analysis of both disaggregated and aggregated VMS message types. In addition, the adjusted odds ratios did not significantly differ from the unadjusted ones, and no apparent departure of impact was detected for both types of VMS messages. However, traffic metrics such as downstream standard deviation of speed and average occupancy, and weather conditions such as rainfall and snowfall were potential risk factors to the crash occurrence according to the regression results.

## **Chapter 6: Conclusion**

The VMS has been broadly used in Minnesota to provide real-time traffic information and guidance to road users since 1960s. Currently, permanent VMS are distributed along many freeways and trunk highways, and are able to be activated anytime needed. To guarantee its proper installation, effective operation, and safe use, the evaluation of VMS is important. In terms of previous studies, the evaluations typically were divided into two parts: qualitative evaluation and quantitative evaluation. Commonly, the former were drivers-survey oriented, and the later were based on analyses of archived or real-time traffic data, or by simulation. However, unlike previous studies, these analyses were conducted under numerous scenarios with highly disaggregated data. In addition, many risk factors other than VMS were identified and controlled.

This study selected five VMS devices along Interstate-94 between downtown Minneapolis and St. Paul. The data collection period was from January 2006 to December 2012. VMS logs were used to determine the activation or the deactivation of the signs, and the types of the sign messages. Those logs were requested from MnDOT. Crash data were extracted from the MnCMAT database. Crash times, locations, types, road surface, and weather conditions were critical information that used in this study. Real-time traffic data were downloaded using MnDOT's data extraction tools. The raw data, which were speed, occupancy, and volume, were used to calculate different traffic metrics. The majority of weather data were collected from the stations of Weather Underground. Precipitation type was primarily utilized in the analyses.

### **6.1 Discussion of Methodologies and Findings**

This study evaluated the performance of VMS into two aspects. In the first aspect, the speed change in the VMS impact regions was investigated. The null hypothesis was that, the vehicle speed in the impact regions did not significantly differ from the initial speed, under the impact of the VMS deployment. In the second aspect, the association between the displaying different types of sign messages, and the probability of crashes was analyzed. The null hypothesis was that, the probability of crashes when the VMS messages were active, was not significantly different from the probability of crashes

during the VMS messages were blank. Both hypotheses were tested in various scenarios by statistical models, with highly disaggregated data.

In Chapter 4, by comparing the vehicles' speed in the VMS impact regions before and after the change in VMS status, results indicated that no difference was detected when VMS was deactivated, but speed decreases were observed when VMS was activated. Normal linear models were used to address the issue in two types of analyses. For the analysis of difference scores, the speed change was assumed zero, and the change was assumed not to vary with initial speed. The response was the change of speed, and the explanatory variables included the times of day for message displaying (AM Peak, PM Peak, and Off Peak), the VMS statuses (activated, and deactivated), and the weather conditions (normal, and adverse weather). The analyses were performed in two scenarios, including eight conditions, with the base condition under the circumstance that, a VMS was deactivated during off-peak hours in normal weather conditions. The overall finding was that VMS was not associated with a notable fluctuation in speed. Another interesting finding was that adverse weather conditions were not likely to affect the change of speed. In addition, drivers tended to slow down within only 2.0 mph in average, when a VMS was activated. However, the drop of speed showed no significant difference during the AM peak hours (1.4 mph), PM peak hours (1.3 mph), and off-peak hours (1.1 mph). Briefly, the deployment of VMS in the selected impact regions had no potentially significant influence on the speed. In other words, drivers did not experience noteworthy fluctuation in speed when driving through the VMS impact regions, of any time of day, and under any weather condition.

An analysis of covariance was conducted after the analysis of difference scores, in order to verify the assumptions that the initial speed did not affect the results. A normal linear model was used as well. However, the response was the impact speed, which was the average speed after the change of VMS status; the explanatory variables included the initial speed, which was the average speed before the change of VMS status; and the other variables used in the difference scores analysis. The regression results revealed that the initial speed and the impact speed were highly correlated, and the coefficient of the initial speed was 0.934, which was close to 1.0. In addition, the patterns of the

coefficients from both models were similar. In summary, the results from analysis of difference scores were accepted from a practical significance standpoint. VMS did not have substantial influence on the speeds in the impact regions.

In Chapter 5, a case-control study was conducted to analyze the association between the deployment of VMS and the crash occurrence. The null hypothesis was that the probability of crashes given VMS was the same with the probability of crashes given no VMS. The analysis revealed that VMS do not cause crashes, as the null hypothesis was not rejected. To test the hypothesis, first, an unadjusted odds ratio was estimated by a 2×2 contingency table, which showed the distribution of crashes and non-crash controls corresponding to the VMS statuses. The estimated odds ratio,  $\hat{\theta}_0 = 1.25$ , was not significantly different from 1.0, at the  $\alpha = 0.05$  significant level.

Second, a logistic regression model was constructed to estimate the adjusted odds ratio, while controlling for other possible crash causal factors, such as downstream traffic metrics and weather conditions. In addition, the VMS site factor and the crash type factor were considered in the model. Moreover, VMS messages were disaggregated into three types. The results from a stepwise model selection found that, the coefficients for VMS sites variables, and crash diagram variables were not significantly different from zero. This indicated that the probability of crashes showed no detectable difference in the selected VMS impact regions, and it did not vary with different crash types. In addition, the VMS was not a significant risk factor with each type of messages.

The adjusted odds ratios were estimated by taking the exponential of coefficients of VMS variables. For “Danger/Warning” message type, the estimated odds ratio was 1.17; and for “Informative/Common” message type, the estimated odds ratio was 1.23. Although both odds ratios were greater than 1.0, there was no evidence to reject the fact that, they were not significantly greater than 1.0. Those results revealed that the probability of crash under the impact of any type of VMS messages showed no significant difference from the probability of crash when a VMS was blank. Thus, in this study, the deployment of VMS was not a potential risk factor to crashes.

## **6.2 Achievements and Limitations**

In this study, many improvements were made compared to previous evaluations. First, the VMS impact region was reasonably clarified, which enabled the crash data and traffic data to be accurately prepared. Second, crash times used in this study were estimated, rather than extracted directly from police reports. Third, the influence of downstream shockwaves was identified and considered when analyzing the speed change under the impact of VMS deployment. Fourth, times of day, and weather conditions were considered, when studying the speed change, which were seldom recognized in the previous evaluations. Last, the adjusted odds ratios were derived with numerous risk variables being controlled, and were estimated for each type of VMS messages. Those achievements are capable to be applied to any safety evaluation of VMS performance in the future.

The definition of VMS impact region is important. It draws the boundaries for crash data collection, and for the selection of loop detectors for traffic data collection. In this study, it was defined as road segment starting from a VMS panel to 860 feet upstream away. The specific distance, 860 feet, was mandated by the Manual of Uniform Traffic Control Devices (MUTCD) as the minimum legible distance. Beyond this distance, drivers might not be able to read the sign, and its effect becomes problematic.

The methodology for estimating crash times is recommended to all the safety evaluation of VMS, if specific traffic data are available. In this study, crash times were estimated based on the times when significant variations of traffic were observed at the adjacent loop detectors close to the crashes, and the shockwave travel time. Crash times recorded by police reports or other relative sources are not reliable, especially for the VMS “on-and-off” evaluation. However, crashes times are required to be linked as precisely as possible to the times when VMS are activated or deactivated. Therefore, evaluation results developed with the direct use of reported crash times may bias the results.

Previous studies tended to overestimate the impact of VMS on speed reduction in their impact regions. One of the major reasons was that the speed reduction was not caused by the drivers reading VMS messages, but by downstream traffic conditions. For example, if



a VMS message indicated there was a crash ahead, traffic slowed down close to the VMS site possibly result from the shockwaves caused by the crash, but not the presence of VMS message. In this study, the cases, whose VMS impact regions could potentially be affected by the downstream shockwaves were identified, and removed from the analyses. The speed change was studied under the normal conditions without disturbance by downstream “unknown events”. The reductions of speed influenced by the activation of VMS in this study were all within 2.0 mph, which were lower than the previous study results (Haghani et al., 2013; and Erke et al., 2007).

The hypotheses, which the impact of VMS on speed change was not influenced by the times of day, and adverse weather conditions, were tested in this study. Previous evaluations tended to assume the hypotheses were true without further investigations. However, this study classified the times of VMS deployment into three ranges, and linked these with real-time weather conditions. Although the results revealed that the speed change did not have significant change during different times of the day, and in adverse weather conditions, it was more comfortable to draw the conclusions with those variables controlled.

Finally, the adjusted odds ratio, which comparing the probabilities of crashes under the presence of VMS messages, was estimated with numerous traffic metrics, and weather conditions identified and controlled in a logistic regression model. Few existing publications consider a comprehensive collection of crash causal factors, when evaluating the safety performance of VMS. Overlooking those factors could result in misjudging the impact of VMS on crash occurrence. For example, a crash caused by heavy snow weather condition should not be attributed to the deployment of VMS. In addition, it is not recommended to aggregate the data collected from different VMS sites, unless the influence of VMS for each site to the crash occurrence is the same with the others. In this study, the VMS sites variables were considered in the regression model, and results indicated there was no significant difference for crash probabilities among the selected sites. Moreover, the odds ratios were estimated for different types of VMS messages, in order to explore their impacts on crash occurrence respectively, although their differences were not significant.

Limitations did exist in this study. First, the study was conducted in Minnesota, so the results might not apply to other states directly, due to the adverse weather conditions in winter. The impact of snow on crash occurrence could possibly affect the odds ratios for VMS deployment. This study did control the weather condition, but the samples of adverse weather were not of a sufficient size. Second, the VMS sites selected in this study were all distributed in one corridor. Since the study period for data collection lasted for seven years, it was challenging to collect all the real-time VMS logs data, crash data, traffic data, and weather data for multiple corridors. In addition, the requirements for the accuracy of data were strict, and it was time consuming to store and refine the raw data. Other corridor studies will be considered in the future. Moreover, the study results cannot be used or compared directly if road curvature exists in the VMS impact region. If so, the curvature should be considered as another control variable in the model, or a curve-adjusted index should be derived. Finally, the models in this study did not capture the influence of upstream traffic conditions to crash occurrence, due to the availability of the data. However, the methodologies stated in this study are recommended for the quantitative safety evaluation of the permanent VMS.

## Bibliography

- Abdel-Aty, Mohamed, Al-Ahad Ekram, Helai Huang, and Keechoo Choi. "A study on crashes related to visibility obstruction due to fog and smoke." *Accident Analysis and Prevention*. 43. (2011): 1730-1737. Print.
- Abdel-Aty, Mohamed, Nizam Uddin, Anurag Pande, M. Fathy Abdalla, and Liang Hsia. "Predicting freeway crashes from loop detector data by matched case-control logistic regression." *Transportation Research Record*. 1897. (2004), 88–95.
- Abdel-Aty, Mohamed, and Pande Anurag. "Identifying Crash Propensity Using Specific Traffic Speed Conditions." *Journal of Safety Research*. 36. (2004): 97-108. Print.
- Abdel-Aty, Mohamed, and Rajashekar Pemmanaboina. "Calibrating a Real-Time Traffic Crash-Prediction Model Using Archived Weather and its Traffic Data." *IEEE Transactions on Intelligent Transportation Systems*. 7.2 (2006): 167-174. Print.
- Agresti, Alan. "Describing Contingency Tables." *Categorical Data Analysis*. New York: Wiley, 1990. 36-59. Print.
- Agresti, Alan. "Inference for Contingency Tables." *Categorical Data Analysis*. New York: Wiley, 1990. 70-102. Print.
- Andreescu, Mircea-Paul, and David B. Frost. "Weather and traffic accidents in Montreal, Canada." *Climate Research Clim Res*. February (1998): 225-230. Print.
- Bakiri, Sarah, Cédric Galéra, Emmanuel Lagarde, Magali Laborey, Benjamin Contrand, Régis Ribéreau-Gayon, Louis-Rachid Salmi, Catherine Gabaude, Alexandra Fort, Bertrand Maury, Céline Lemerrier, Maurice Cours, Manuel-Pierre Bouvard, and Ludivine Orriols. "Distraction and Driving: Results from a Case-control Responsibility Study of Traffic Crash Injured Drivers Interviewed at the Emergency Room." *Accident Analysis & Prevention* 59 (2013): 588-592. Web.
- Chang, Gang-Len, Xiang Fei, and Jean Yves. "Interrelations between Variable Message Signs and Detour Operations in the I-95 Corridor." *Maryland State Highway Administration Office of Policy & Research*, Aug. 2002. Web.  
<[http://roads.maryland.gov/OPR\\_Research/MD-02-SP007B4G-VMS-and-Detour-Operations-in-the-I-95-Corridor-Report.pdf](http://roads.maryland.gov/OPR_Research/MD-02-SP007B4G-VMS-and-Detour-Operations-in-the-I-95-Corridor-Report.pdf)>.

- Chen, Shaokuan, Mingjun Liu, et al. "Effects of Variable Message Signs (VMS) for Improving Congestions." *IEEE Publishing*, 2008. 416-419. Web.
- Christoforou, Zoi, Simon Cohen, and Matthew G. Karlaftis. "Identifying Crash Type Propensity Using Real-time Traffic Data on Freeways." *Journal of Safety Research* 42.1 (2011): 43-50. Web.
- Chung, Younshik. "Assessment of Non-recurrent Traffic Congestion Caused by Freeway Work Zones and Its Statistical Analysis with Unobserved Heterogeneity." *Transport Policy* 18.4 (2011): 587-594. Web.
- Dobson, Annette J., and Adrian G. Barnett. "Model Fitting." *An Introduction to Generalized Linear Models*. Boca Raton: Chapman & Hall, 2008. 19-40. Print.
- Dobson, Annette J., and Adrian G. Barnett. "Normal Linear Models." *An Introduction to Generalized Linear Models*. Boca Raton: Chapman & Hall, 2008. 89-118. Print.
- Eisenberg, Daniel, and Kenneth E. Warner. "Effects of Snowfalls on Motor Vehicle Collisions, Injuries, and Fatalities." *American Journal of Public Health*. 95(1). January (2005). Print.
- Elefteriadou, Lily. "Mathematical and Empirical Models." *An Introduction to Traffic Flow Theory*. New York: Springer, 2014. 129-135. Print.
- Erke, A., Sagberg, F., Hagman, R. "Effects of route guidance variable message signs (VMS) on driver behavior." *Institute of Transport Economics*. Gaustadalleen 21, NO-0349 Oslo, Norway, 2007.
- "Frequently Asked Questions." *FHWA Road Weather Management*. Web.  
<<http://ops.fhwa.dot.gov/weather/faq.htm>>.
- Haghani, Ali, Masoud Masoud, Robin Fish, and Azadeh Nouruzi. "Evaluation of Dynamic Message Signs and Their Potential Impact on Traffic Flow." *Maryland State Highway Administration Office of Policy & Research*, Apr. 2013. Web.  
<[http://ntl.bts.gov/lib/47000/47400/47478/MD-13-SP109B4C\\_Impact-of-DMS-Messages\\_Report.pdf](http://ntl.bts.gov/lib/47000/47400/47478/MD-13-SP109B4C_Impact-of-DMS-Messages_Report.pdf)>.
- Harder, Kathleen A., John Bloomfield, and Benjamin J. Chihak. "The Effectiveness and Safety of Traffic and Non-Traffic Related Messages Presented on Changeable

- Message Signs (CMS)" *Minnesota Department of Transportation Office of Research Services*, June 2003. Web. <<http://www.lrrb.org/PDF/200427.pdf>>.
- Hosmer, David W., and Stanley Lemeshow. *Applied Logistic Regression*. New York: Wiley, 2000. Print.
- Hourdos, John N, Vishnu Garg, Panos G Michalopoulos, and Gary A Davis. "Real-Time Detection of Crash-Prone Conditions at Freeway High-Crash Locations." *Journal of the Transportation Research Board*. 1968 February (2007): 83-91. Print.
- Hoye, Alena, Michael Sorensen, Rune Elvik, Juned Akhtar, Tor-Olav Nævestad, and Truls Vaa. "Summary: Evaluation of variable message signs in Trondheim." *Institute of Transport Economics Norwegian Centre for Transport Research*. 2011. Web.<<https://www.toi.no/getfile.php/Publikasjoner/T%C3%98I%20rapporter/2011/1153-2011/1153-2011-sum.pdf>>.
- Kallbekken, Steffen, and Marianne Aasen. "The demand for earmarking: Results from a focus group study." *ScienceDirect Journals*. 69.11 (2010): 2183-2190. Print.
- Kolisetty, K Ganesh Babu, Takamasa Iryo, Yasuo Asakura, and Katsuhiko Kuroda. "Effect of Variable Message Signs on Driver Speed Behavior on a Section of Expressway under Adverse Fog Conditions - A Driving Simulator Approach." *Journal of Advanced Transportation*. 40.1 (2005): 47-74. Print.
- Lee, Chris, Bruce Hellinga, and Frank Saccomanno. "Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic." *Journal of the Transportation Research Board* 1840 (2003): 67–77.
- Lee, Chris, Frank Saccomanno, Bruce Hellinga. "Analysis of Crash Precursors on Instrumented Freeways." *Transportation Research Record: Journal of the Transportation Research Board* 1784 (2002): 1–8.
- Lee, Dongmin, Chungwon Lee, and Martin T. Pietrucha. "Evaluation of Driver Satisfaction of Travel Information on Variable Message Signs Using Fuzzy Aggregation." *Journal of Advanced Transportation*. 42. February (2007): 5-22. Print.

- Levinson, D and Hong Huo "Effectiveness of VMS Using Empirical Loop Detector Data." *Presented at Transportation Research Board Conference*, January 12 – 16 2003 Washington DC (Session 565)
- Mende, I. *An Analysis of Snowstorm-related Accidents in Metropolitan Toronto*. Thesis. University of Toronto, 1982. Department of Civil Engineering, Print.
- "Mn/DOT Data Tools". <<http://data.dot.state.mn.us/datatools/>>.
- Mounce, John M., Gerald Ullman, Geza Pesti, and Valmon Pezoldt. "Guidelines for the Evaluation of Dynamic Message Sign Performance" *Texas Department of Transportation Research and Technology Implementation Office*, Mar. 2007. Web. <<http://tti.tamu.edu/documents/0-4772-1.pdf>>.
- Oh, Jun-Seok, Cheol Oh, Stephen G. Ritchie, and Myungsoon Chang. "Real-time Estimation of Freeway Accident Likelihood." *In: Proceedings of the 80th Annual Meeting of the Transportation Research Board*, Washington, DC. (2001).
- Pisano, Paul A., Lynette C. Goodwin, and Michael A. Rossetti. "U.S. Highway Crashes in Adverse Road Weather Conditions." *Federal Highway Administration (FHWA)*, 2008. Web. 11 Aug. 2013. <<http://ams.confex.com/ams/pdfpapers/133554.pdf>>.
- Quddus, Mohammed A., Chao Wang, and Stephen G. Ison "Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models." *Journal of Transportation Engineering-ASCE*. 136(5). (2010): 424-435. Print.
- Ran, Bin, Bridget Barrett, and Emily Johnson. "Evaluation of Variable Message Signs in Wisconsin: Driver Survey." Wisconsin Department of Transportation, Oct. 2004. Web. <<http://wisdotresearch.wi.gov/wp-content/uploads/45-17variablemessagesigns1.pdf>>.
- Roess, Roger P., Elena S. Prassas, and William R. McShane. "Traffic Stream Characteristics." *Traffic Engineering*. 3rd ed. Upper Saddle River, NJ: Pearson/Prentice Hall, 2004. 105-118. Print.
- Roshandeh, A.M., and O.C. Puan. "Assessment of impact of variable message signs on traffic surveillance in Kuala Lumpur." *IEEE International Conference on Intelligence and Security Informatics*. 2009. 223-225. Web.

- Schlesselman, James J., and Paul D. Stolley. *Case-control Studies: Design, Conduct, Analysis*. New York: Oxford UP, 1982. Print.
- Schroeder, Jeremy L., and Michael J. Demetsky. "Data Analysis of Diversion and Dynamic Message Signs." *Journal of Transportation Engineering* 137.9 (2011): 608-14. Web.
- Shang, Huayan, and Huapu Lu. "Impact of VMS Route Recommendation on Road Network." *IEEE Publishing 2009*. 169-173. Print.
- Shively, Thomas S., Kara Kockelman, and Paul Damien. "A Bayesian semi-parametric model to estimate relationships between crash counts and roadway characteristics." *Transportation Research Part B*. 44. (2010): 699-715. Print.
- Strong, Christopher, Zhirui Ye, and Xianming Shi. "Safety Effects of Winter Weather: The State of Knowledge and Remaining Challenges." *Transport Reviews*. 30.November (2010): 677-699. Print.
- Tay, Richard, and Alexandre G. De Barros. "Public perceptions of the use of dynamic message signs." *Journal of Advanced Transportation*. 42(1). (2008): 95-110. Web.
- Turnbull, Katie and Kevin Balke. "Minnesota Urban Partnership Agreement National Evaluation: Safety Data Test Plan." *U.S. Department of Transportation Research and Innovative Technology Administration*. Web. Nov. 2009.  
<<http://www.upa.dot.gov/docs/fhwajpo10012/index.htm>>
- "Twin Cities Climate Data." *Minnesota Climatology Working Group*.  
<<http://climate.umn.edu/>>.
- U.S. Minnesota Department of Transportation. *2012 CMS Manual of Practice*. By Iteris, Inc. Minnesota Department of Transportation Research Services, Nov. 2012. Web. <<http://dotapp7.dot.state.mn.us/edms/download?docId=1244587>>.
- U.S. Minnesota Department of Transportation. *Guidelines for Changeable Message Sign (CMS) Use*. Sept. 2000. Web. <<http://www.dot.state.mn.us/trafficeng/publ/cms-use/cms-use-guidelines.pdf>>.

- U.S. Minnesota Department of Transportation. Regional Transportation Management Center. *Twin Cities Metropolitan Area Freeway and Major Arterials Crash Summary*. 2011. Print.
- Wardman, M, P.W. Bonsall, and J.D. Shires. "Driver Response to Variable Message Signs: A Stated Preference Investigation." *Transportation Research Part C-emerging Technologies*. (1997): 389-405. Web
- Wu, Xinkai, and Henry X. Liu. "A Shockwave Profile Model for Traffic Flow on Congested Urban Arterials." *Transportation Research Part B: Methodological* 45.10 (2011): 1768-1786. Web.
- Xu, Chengcheng, Wei Wang, and Pan Liu. "Identifying crash-prone traffic conditions under different weather on freeways." *Journal of Safety Research*. 46. (2013): 135-144. Print.
- Yeo, Hwasoo, Kitae Jang, et al. "Impact of traffic states on freeway crash involvement rates." *Accident Analysis and Prevention*. 50. (2013): 713-723. Print.
- Yu, Rongjie, Mohamed Abdel-Aty, and Mohamed Ahmed. "Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors." *Accident Analysis and Prevention*. 50. (2013): 371-376. Print.
- Yu, Rongjie, and Mohamed Abdel-Aty. "Investigating the Different Characteristics of Weekday and Weekend Crashes." *Journal of Safety Research*. 46. (2013): 91-97. Print.
- Zheng Zuduo, Soyoung Ahn, and Christopher.M Monsere. "Impact of traffic oscillations on freeway crash occurrences." *Accident Analysis and Prevention*. 42. (2010): 626-636. Print.
- Zheng, Zuduo. "Empirical Analysis on Relationship between Traffic Conditions and Crash Occurrences." *Social and Behavioral Sciences*. 43. (2012): 302-312. Print.